

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/53, 15/55, 15/60, C12P 7/18, C12N 9/04, 9/16, 9/88	A1	(11) International Publication Number: WO 98/21339 (43) International Publication Date: 22 May 1998 (22.05.98)
(21) International Application Number: PCT/US97/20292 (22) International Filing Date: 10 November 1997 (10.11.97) (30) Priority Data: 60/030,601 13 November 1996 (13.11.96) US (71) Applicants (for all designated States except US): E.I. DU PONT DE NEMOURS AND COMPANY [US/US]; 1007 Market Street, Wilmington, DE 19898 (US). GENENCOR INTERNATIONAL, INC. [US/US]; 4 Cambridge Place, 1870 Winton Road, Rochester, NY 14618 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): GATENBY, Anthony, Arthur [US/US]; 2309 Baynard Boulevard, Wilmington, DE 19802 (US). HAYNIE, Sharon, Loretta [US/US]; 963 North Randolph Street, Philadelphia, PA 19123 (US). NAGARAJAN, Vasanth [US/US]; 13 Dickinson Lane, Wilmington, DE 19807 (US). NAIR, Ramesh, V. [US/US]; 905-D Cloister Road, Wilmington, DE 19809 (US). NAKAMURA, Charles, E. [US/US]; 2 Mount Vernon Drive, Claymont, DE 19703 (US). PAYNE, Mark, Scott [US/US]; 2408 Newell Drive, Wilmington, DE 19808 (US). PICATAGGIO, Stephen, Kenneth [US/US]; 17 Meadow Wood Lane, Landenberg, PA 19350 (US). DIAS-TORRES, Maria [ES/US];		58 N. El Camino Real #118, San Mateo, CA 94401 (US). HSU, Amy, Kuang-Hua [CN/US]; 528 Keelson Circle, Redwood City, CA 94065 (US). LAREAU, Richard, D. [US/US]; 197A Easy Street, Mountain View, CA 94043 (US). TRIMBUR, Donald, E. [US/US]; 349 Orchard Avenue, Redwood City, CA 94601 (US). WHITED, Gregory, M. [US/US]; 304 South Road, Belmont, CA 94002 (US). (74) Agent: FLOYD, Linda, A.; E.I. du Pont de Nemours and Company, Legal Patent Records Center, 1007 Market Street, Wilmington, DE 19898 (US). (81) Designated States: AU, BR, CA, CN, ID, IL, JP, KR, MX, SG, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(54) Title: METHOD FOR THE PRODUCTION OF 1,3-PROPANEDIOL BY RECOMBINANT ORGANISMS (57) Abstract Recombinant organisms are provided comprising genes encoding glycerol-3-phosphate dehydrogenase, glycerol-3-phosphatase, glycerol dehydratase and 1,3-propanediol oxidoreductase activities useful for the production of 1,3-propanediol from a variety of carbon substrates.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

TITLEMETHOD FOR THE PRODUCTION OF 1,3-PROPANEDIOL
BY RECOMBINANT ORGANISMSFIELD OF INVENTION

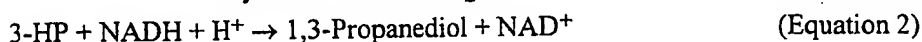
5 The present invention relates to the field of molecular biology and the use of recombinant organisms for the production of desired compounds. More specifically it describes the expression of cloned genes for glycerol-3-phosphate dehydrogenase (G3PDH) and glycerol-3-phosphatase (G3P phosphatase), glycerol dehydratase (*dhaB*), and 1,3-propanediol oxidoreductase (*dhaT*), either separately
10 or together, for the enhanced production of 1,3-propanediol.

BACKGROUND

1,3-Propanediol is a monomer having potential utility in the production of polyester fibers and the manufacture of polyurethanes and cyclic compounds.

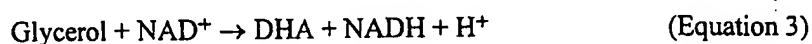
A variety of chemical routes to 1,3-propanediol are known. For example
15 ethylene oxide may be converted to 1,3-propanediol over a catalyst in the presence of phosphine, water, carbon monoxide, hydrogen and an acid, by the catalytic solution phase hydration of acrolein followed by reduction, or from hydrocarbons such as glycerol, reacted in the presence of carbon monoxide and hydrogen over catalysts having atoms from group VIII of the periodic table. Although it is
20 possible to generate 1,3-propanediol by these methods, they are expensive and generate waste streams containing environmental pollutants.

It has been known for over a century that 1,3-propanediol can be produced from the fermentation of glycerol. Bacterial strains able to produce 1,3-propanediol have been found, for example, in the groups *Citrobacter*, *Clostridium*,
25 *Enterobacter*, *Ilyobacter*, *Klebsiella*, *Lactobacillus*, and *Pelobacter*. In each case studied, glycerol is converted to 1,3-propanediol in a two step, enzyme catalyzed reaction sequence. In the first step, a dehydratase catalyzes the conversion of glycerol to 3-hydroxypropionaldehyde (3-HP) and water (Equation 1). In the second step, 3-HP is reduced to 1,3-propanediol by a NAD⁺-linked
30 oxidoreductase (Equation 2).



35 The 1,3-propanediol is not metabolized further and, as a result, accumulates in high concentration in the media. The overall reaction consumes a reducing equivalent in the form of a cofactor, reduced β -nicotinamide adenine dinucleotide (NADH), which is oxidized to nicotinamide adenine dinucleotide (NAD⁺).

The production of 1,3-propanediol from glycerol is generally performed under anaerobic conditions using glycerol as the sole carbon source and in the absence of other exogenous reducing equivalent acceptors. Under these conditions, in for example, strains of *Citrobacter*, *Clostridium*, and *Klebsiella*, a parallel pathway for glycerol operates which first involves oxidation of glycerol to dihydroxyacetone (DHA) by a NAD⁺- (or NADP⁺-) linked glycerol dehydrogenase (Equation 3). The DHA, following phosphorylation to dihydroxyacetone phosphate (DHAP) by a DHA kinase (Equation 4), becomes available for biosynthesis and for supporting ATP generation via, for example, glycolysis.



In contrast to the 1,3-propanediol pathway, this pathway may provide carbon and energy to the cell and produces rather than consumes NADH.

In *Klebsiella pneumoniae* and *Citrobacter freundii*, the genes encoding the functionally linked activities of glycerol dehydratase (*dhaB*), 1,3-propanediol oxidoreductase (*dhaT*), glycerol dehydrogenase (*dhaD*), and dihydroxyacetone kinase (*dhaK*) are encompassed by the *dha* regulon. The *dha* regulons from *Citrobacter* and *Klebsiella* have been expressed in *Escherichia coli* and have been shown to convert glycerol to 1,3-propanediol.

Biological processes for the preparation of glycerol are known. The overwhelming majority of glycerol producers are yeasts, but some bacteria, other fungi and algae are also known to produce glycerol. Both bacteria and yeasts produce glycerol by converting glucose or other carbohydrates through the fructose-1,6-bisphosphate pathway in glycolysis or by the Embden Meyerhof Parnas pathway, whereas, certain algae convert dissolved carbon dioxide or bicarbonate in the chloroplasts into the 3-carbon intermediates of the Calvin cycle. In a series of steps, the 3-carbon intermediate, phosphoglyceric acid, is converted to glyceraldehyde 3-phosphate which can be readily interconverted to its keto isomer dihydroxyacetone phosphate and ultimately to glycerol.

Specifically, the bacteria *Bacillus licheniformis* and *Lactobacillus lycopersica* synthesize glycerol, and glycerol production is found in the halotolerant algae *Dunaliella sp.* and *Asteromonas gracilis* for protection against high external salt concentrations (Ben-Amotz et al., *Experientia* 38, 49-52, (1982)). Similarly, various osmotolerant yeasts synthesize glycerol as a protective measure. Most strains of *Saccharomyces* produce some glycerol during alcoholic fermentation, and this can be increased physiologically by the application of

osmotic stress (Albertyn et al., *Mol. Cell. Biol.* 14, 4135-4144, (1994)). Earlier this century commercial glycerol production was achieved by the use of *Saccharomyces* cultures to which "steering reagents" were added such as sulfites or alkalis. Through the formation of an inactive complex, the steering agents
5 block or inhibit the conversion of acetaldehyde to ethanol; thus, excess reducing equivalents (NADH) are available to or "steered" towards DHAP for reduction to produce glycerol. This method is limited by the partial inhibition of yeast growth that is due to the sulfites. This limitation can be partially overcome by the use of alkalis which create excess NADH equivalents by a different mechanism. In this
10 practice, the alkalis initiated a Cannizzaro disproportionation to yield ethanol and acetic acid from two equivalents of acetaldehyde.

The gene encoding glycerol-3-phosphate dehydrogenase (DAR1, GPD1) has been cloned and sequenced from *S. diastaticus* (Wang et al., *J. Bact.* 176, 7091-7095, (1994)). The DAR1 gene was cloned into a shuttle vector and used to
15 transform *E. coli* where expression produced active enzyme. Wang et al. (supra) recognize that DAR1 is regulated by the cellular osmotic environment but do not suggest how the gene might be used to enhance 1,3-propanediol production in a recombinant organism.

Other glycerol-3-phosphate dehydrogenase enzymes have been isolated:
20 for example, sn-glycerol-3-phosphate dehydrogenase has been cloned and sequenced from *S. cerevisiae* (Larason et al., *Mol. Microbiol.* 10, 1101, (1993)) and Albertyn et al., (*Mol. Cell. Biol.* 14, 4135, (1994)) teach the cloning of GPD1 encoding a glycerol-3-phosphate dehydrogenase from *S. cerevisiae*. Like Wang et al. (supra), both Albertyn et al. and Larason et al. recognize the osmo-sensitivity
25 of the regulation of this gene but do not suggest how the gene might be used in the production of 1,3-propanediol in a recombinant organism.

As with G3PDH, glycerol-3-phosphatase has been isolated from *Saccharomyces cerevisiae* and the protein identified as being encoded by the GPP1 and GPP2 genes (Norbeck et al., *J. Biol. Chem.* 271, 13875, (1996)). Like
30 the genes encoding G3PDH, it appears that GPP2 is osmosensitive.

Although biological methods of both glycerol and 1,3-propanediol production are known, it has never been demonstrated that the entire process can be accomplished by a single recombinant organism.

Neither the chemical nor biological methods described above for the
35 production of 1,3-propanediol are well suited for industrial scale production since the chemical processes are energy intensive and the biological processes require the expensive starting material, glycerol. A method requiring low energy input and an inexpensive starting material is needed. A more desirable process would

incorporate a microorganism that would have the ability to convert basic carbon sources such as carbohydrates or sugars to the desired 1,3-propanediol end-product.

Although a single organism conversion of fermentable carbon source other than glycerol or dihydroxyacetone to 1,3-propanediol would be desirable, it has been documented that there are significant difficulties to overcome in such an endeavor. For example, Gottschalk et al. (EP 373 230) teach that the growth of most strains useful for the production of 1,3-propanediol, including *Citrobacter freundii*, *Clostridium autobutylicum*, *Clostridium butylicum*, and *Klebsiella pneumoniae*, is disturbed by the presence of a hydrogen donor such as fructose or glucose. Strains of *Lactobacillus brevis* and *Lactobacillus buchneri*, which produce 1,3-propanediol in co-fermentations of glycerol and fructose or glucose, do not grow when glycerol is provided as the sole carbon source, and, although it has been shown that resting cells can metabolize glucose or fructose, they do not produce 1,3-propanediol. (Veiga DA Cunha et al., *J. Bacteriol.* 174, 1013 (1992)). Similarly, it has been shown that a strain of *Ilyobacter polytropus*, which produces 1,3-propanediol when glycerol and acetate are provided, will not produce 1,3-propanediol from carbon substrates other than glycerol, including fructose and glucose. (Steib et al., *Arch. Microbiol.* 140, 139 (1984)). Finally Tong et al. (*Appl. Biochem. Biotech.* 34, 149 (1992)) has taught that recombinant *Escherichia coli* transformed with the *dha* regulon encoding glycerol dehydratase does not produce 1,3-propanediol from either glucose or xylose in the absence of exogenous glycerol.

Attempts to improve the yield of 1,3-propanediol from glycerol have been reported where co-substrates capable of providing reducing equivalents, typically fermentable sugars, are included in the process. Improvements in yield have been claimed for resting cells of *Citrobacter freundii* and *Klebsiella pneumoniae* DSM 4270 cofermenting glycerol and glucose (Gottschalk et al., *supra.*, and Tran-Dinh et al., DE 3734 764); but not for growing cells of *Klebsiella pneumoniae* ATCC 25955 cofermenting glycerol and glucose, which produced no 1,3-propanediol (I-T. Tong, Ph.D. Thesis, University of Wisconsin-Madison (1992)). Increased yields have been reported for the cofermentation of glycerol and glucose or fructose by a recombinant *Escherichia coli*; however, no 1,3-propanediol is produced in the absence of glycerol (Tong et al., *supra.*). In these systems, single organisms use the carbohydrate as a source of generating NADH while providing energy and carbon for cell maintenance or growth. These disclosures suggest that sugars do not enter the carbon stream that produces 1,3-propanediol. In no case is 1,3-propanediol produced in the absence of an

exogenous source of glycerol. Thus the weight of literature clearly suggests that the production of 1,3-propanediol from a carbohydrate source by a single organism is not possible.

The problem to be solved by the present invention is the biological production of 1,3-propanediol by a single recombinant organism from an inexpensive carbon substrate such as glucose or other sugars. The biological production of 1,3-propanediol requires glycerol as a substrate for a two step sequential reaction in which a dehydratase enzyme (typically a coenzyme B₁₂-dependent dehydratase) converts glycerol to an intermediate, 3-hydroxypropionaldehyde, which is then reduced to 1,3-propanediol by a NADH- (or NADPH) dependent oxidoreductase. The complexity of the cofactor requirements necessitates the use of a whole cell catalyst for an industrial process which utilizes this reaction sequence for the production of 1,3-propanediol. Furthermore, in order to make the process economically viable, a less expensive feedstock than glycerol or dihydroxyacetone is needed. Glucose and other carbohydrates are suitable substrates, but, as discussed above, are known to interfere with 1,3-propanediol production. As a result no single organism has been shown to convert glucose to 1,3-propanediol.

Applicants have solved the stated problem and the present invention provides for bioconverting a fermentable carbon source directly to 1,3-propanediol using a single recombinant organism. Glucose is used as a model substrate and the bioconversion is applicable to any existing microorganism. Microorganisms harboring the genes encoding glycerol-3-phosphate dehydrogenase (G3PDH), glycerol-3-phosphatase (G3P phosphatase), glycerol dehydratase (*dhaB*), and 1,3-propanediol oxidoreductase (*dhaT*), are able to convert glucose and other sugars through the glycerol degradation pathway to 1,3-propanediol with good yields and selectivities. Furthermore, the present invention may be generally applied to include any carbon substrate that is readily converted to 1) glycerol, 2) dihydroxyacetone, or 3) C₃ compounds at the oxidation state of glycerol (e.g., glycerol 3-phosphate) or 4) C₃ compounds at the oxidation state of dihydroxyacetone (e.g., dihydroxyacetone phosphate or glyceraldehyde 3-phosphate).

SUMMARY OF THE INVENTION

The present invention provides a method for the production of 1,3-propanediol from a recombinant organism comprising:

- (i) transforming a suitable host organism with a transformation cassette comprising at least one of (a) a gene encoding a glycerol-3-phosphate dehydrogenase activity; (b) a gene encoding a glycerol-3 phosphatase activity;

- (c) genes encoding a dehydratase activity; and (d) a gene encoding 1,3-propanediol oxidoreductase activity, provided that if the transformation cassette comprises less than all the genes of (a)-(d), then the suitable host organism comprises endogenous genes whereby the resulting transformed host organism comprises at least one of each of genes (a)-(d);
- (ii) culturing the transformed host organism under suitable conditions in the presence of at least one carbon source selected from the group consisting of monosaccharides, oligosaccharides, polysaccharides, or a one carbon substrate whereby 1,3-propanediol is produced; and
- (iii) recovering the 1,3-propanediol.

The invention further provides transformed hosts comprising expression cassettes capable of expressing glycerol-3-phosphate dehydrogenase, glycerol-3-phosphatase, glycerol dehydratase and 1,3-propanediol oxidoreductase activities for the production of 1,3-propanediol.

- The suitable host organism used in the method is selected from the group consisting of bacteria, yeast, and filamentous fungi. The suitable host organism is more particularly selected from the group of genera consisting of *Citrobacter*, *Enterobacter*, *Clostridium*, *Klebsiella*, *Aerobacter*, *Lactobacillus*, *Aspergillus*, *Saccharomyces*, *Schizosaccharomyces*, *Zygosaccharomyces*, *Pichia*, *Kluyveromyces*, *Candida*, *Hansenula*, *Debaryomyces*, *Mucor*, *Torulopsis*, *Methylobacter*, *Escherichia*, *Salmonella*, *Bacillus*, *Streptomyces* and *Pseudomonas*. Most particularly, the suitable host organism is selected from the group consisting of *E. coli*, *Klebsiella spp.*, and *Saccharomyces spp.* Particular transformed host organisms used in the method are 1) a *Saccharomyces spp.* transformed with a transformation cassette comprising the genes *dhaB1*, *dhaB2*, *dhaB3*, and *dhaT*, wherein the genes are stably integrated into the *Saccharomyces spp.* genome; and 2) a *Klebsiella spp.* transformed with a transformation cassette comprising the genes GPD1 and GPD2;

The preferred carbon source of the invention is glucose.

- The method further uses the gene encoding a glycerol-3-phosphate dehydrogenase enzyme selected from the group consisting of genes corresponding to amino acid sequences given in SEQ ID NO:11, in SEQ ID NO:12, and in SEQ ID NO:13, the amino acid sequences encompassing amino acid substitutions, deletions or additions that do not alter the function of the glycerol-3-phosphate dehydrogenase enzyme. The method also uses the gene encoding a glycerol-3-phosphatase enzyme selected from the group consisting of genes corresponding to amino acid sequences given in SEQ ID NO:33 and in SEQ ID NO:17, the amino acid sequences encompassing amino acid substitutions, deletions or additions that

do not alter the function of the glycerol-3-phosphatase enzyme. The method also uses the gene encoding a glycerol kinase enzyme that corresponds to an amino acid sequence given in SEQ ID NO:18, the amino acid sequence encompassing amino acid substitutions, deletions or additions that do not alter the function of the glycerol kinase enzyme. The method also uses the genes encoding a dehydratase enzyme comprise dhaB1, dhaB2 and dhaB3, the genes corresponding respectively to amino acid sequences given in SEQ ID NO:34, SEQ ID NO:35, and SEQ ID NO:36, the amino acid sequences encompassing amino acid substitutions, deletions or additions that do not alter the function of the dehydratase enzyme.

The method also uses the gene encoding a 1,3-propanediol oxidoreductase enzyme that corresponds to an amino acid sequence given in SEQ ID NO:37, the amino acid sequence encompassing amino acid substitutions, deletions or additions that do not alter the function of the 1,3-propanediol oxidoreductase enzyme.

The invention is also embodied in a transformed host cell comprising:

- (a) a group of genes comprising
 - (1) a gene encoding a glycerol-3-phosphate dehydrogenase enzyme corresponding to the amino acid sequence given in SEQ ID NO:11;
 - (2) a gene encoding a glycerol-3-phosphatase enzyme corresponding to the amino acid sequence given in SEQ ID NO:17;
 - (3) a gene encoding the α subunit of the glycerol dehydratase enzyme corresponding to the amino acid sequence given in SEQ ID NO:34;
 - (4) a gene encoding the β subunit of the glycerol dehydratase enzyme corresponding to the amino acid sequence given in SEQ ID NO:35;
 - (5) a gene encoding the γ subunit of the glycerol dehydratase enzyme corresponding to the amino acid sequence given in SEQ ID NO:36; and
 - (6) a gene encoding the 1,3-propanediol oxidoreductase enzyme corresponding to the amino acid sequence given in SEQ ID NO:37,
 the respective amino acid sequences of (a)(1)-(6) encompassing amino acid substitutions, deletions, or additions that do not alter the function of the enzymes of genes (1)-(6), and
- (b) a host cell transformed with the group of genes of (a), whereby the transformed host cell produces 1,3-propanediol on at least one substrate selected from the group consisting of monosaccharides, oligosaccharides, and polysaccharides or from a one-carbon substrate.

BRIEF DESCRIPTION OF BIOLOGICAL DEPOSITS AND SEQUENCE LISTING

The transformed *E. coli* W2042 (comprising the *E. coli* host W1485 and plasmids pDT20 and pAH42) containing the genes encoding glycerol-3-phosphate

dehydrogenase (G3PDH) and glycerol-3-phosphatase (G3P phosphatase), glycerol dehydratase (*dhaB*), and 1,3-propanediol oxidoreductase (*dhaT*) was deposited on 26 September 1996 with the ATCC under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for the Purpose of Patent Procedure and is designated as ATCC 98188.

S. cerevisiae YPH500 harboring plasmids pMCK10, pMCK17, pMCK30 and pMCK35 containing genes encoding glycerol-3-phosphate dehydrogenase (G3PDH) and glycerol-3-phosphatase (G3P phosphatase), glycerol dehydratase (*dhaB*), and 1,3-propanediol oxidoreductase (*dhaT*) was deposited on 26 September 1996 with the ATCC under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for the Purpose of Patent Procedure and is designated as ATCC 74392.

"ATCC" refers to the American Type Culture Collection international depository located at 12301 Parklawn Drive, Rockville, MD 20852 U.S.A. The designations refer to the accession number of the deposited material.

Applicants have provided 49 sequences in conformity with Rules for the Standard Representation of Nucleotide and Amino Acid Sequences in Patent Applications (Annexes I and II to the Decision of the President of the EPO, published in Supplement No. 2 to OJ EPO, 12/1992) and with 37 C.F.R. 1.821-1.825 and Appendices A and B (Requirements for Application Disclosures Containing Nucleotides and/or Amino Acid Sequences).

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides a method for a biological production of 1,3-propanediol from a fermentable carbon source in a single recombinant organism. The method incorporates a microorganism containing genes encoding glycerol-3-phosphate dehydrogenase (G3PDH), glycerol-3-phosphatase (G3P phosphatase), glycerol dehydratase (*dhaB*), and 1,3-propanediol oxidoreductase (*dhaT*). The recombinant microorganism is contacted with a carbon substrate and 1,3-propanediol is isolated from the growth media.

The present method provides a rapid, inexpensive and environmentally responsible source of 1,3-propanediol monomer useful in the production of polyesters and other polymers.

The following definitions are to be used to interpret the claims and specification.

The terms "glycerol dehydratase" or "dehydratase enzyme" refer to the polypeptide(s) responsible for an enzyme activity that is capable of isomerizing or converting a glycerol molecule to the product 3-hydroxypropionaldehyde. For the purposes of the present invention the dehydratase enzymes include a glycerol

dehydratase (GenBank U09771, U30903) and a diol dehydratase (GenBank D45071) having preferred substrates of glycerol and 1,2-propanediol, respectively. Glycerol dehydratase of *K. pneumoniae* ATCC 25955 is encoded by the genes *dhaB1*, *dhaB2*, and *dhaB3* identified as SEQ ID NOS:1, 2 and 3, respectively.

- 5 The *dhaB1*, *dhaB2*, and *dhaB3* genes code for the α , β , and γ subunits of the glycerol dehydratase enzyme, respectively.

The terms "oxidoreductase" or "1,3-propanediol oxidoreductase" refer to the polypeptide(s) responsible for an enzyme activity that is capable of catalyzing the reduction of 3-hydroxypropionaldehyde to 1,3-propanediol. 1,3-Propanediol
10 oxidoreductase includes, for example, the polypeptide encoded by the *dhaT* gene (GenBank U09771, U30903) and is identified as SEQ ID NO:4.

The terms "glycerol-3-phosphate dehydrogenase" or "G3PDH" refer to the polypeptide(s) responsible for an enzyme activity capable of catalyzing the conversion of dihydroxyacetone phosphate (DHAP) to glycerol-3-phosphate
15 (G3P). *In vivo* G3PDH may be NADH-, NADPH-, or FAD-dependent. Examples of this enzyme activity include the following: NADH-dependent enzymes (EC 1.1.1.8) are encoded by several genes including GPD1 (GenBank Z74071x2) or GPD2 (GenBank Z35169x1) or GPD3 (GenBank G984182) or DAR1 (GenBank Z74071x2); a NADPH-dependent enzyme (EC 1.1.1.94) is encoded by
20 *gpsA* (GenBank U32164, G466746 (cds 197911-196892), and L45246); and FAD-dependent enzymes (EC 1.1.99.5) are encoded by GUT2 (GenBank Z47047x23) or glpD (GenBank G147838) or glpABC (GenBank M20938).

The terms "glycerol-3-phosphatase" or "sn-glycerol-3-phosphatase" or "d,l-glycerol phosphatase" or "G3P phosphatase" refer to the polypeptide(s)
25 responsible for an enzyme activity that is capable of catalyzing the conversion of glycerol-3-phosphate to glycerol. G3P phosphatase includes, for example, the polypeptides encoded by GPP1 (GenBank Z47047x125) or GPP2 (GenBank U18813x11).

The term "glycerol kinase" refers to the polypeptide(s) responsible for an
30 enzyme activity capable of catalyzing the conversion of glycerol to glycerol-3-phosphate or glycerol-3-phosphate to glycerol, depending on reaction conditions. Glycerol kinase includes, for example, the polypeptide encoded by GUT1 (GenBank U11583x19).

The terms "GPD1", "DAR1", "OSG1", "D2830", and "YDL022W" will
35 be used interchangeably and refer to a gene that encodes a cytosolic glycerol-3-phosphate dehydrogenase and characterized by the base sequence given as SEQ ID NO:5.

The term "GPD2" refers to a gene that encodes a cytosolic glycerol-3-phosphate dehydrogenase and characterized by the base sequence given as SEQ ID NO:6.

5 The terms "GUT2" and "YIL155C" are used interchangeably and refer to a gene that encodes a mitochondrial glycerol-3-phosphate dehydrogenase and characterized by the base sequence given in SEQ ID NO:7.

The terms "GPP1", "RHR2" and "YIL053W" are used interchangeably and refer to a gene that encodes a cytosolic glycerol-3-phosphatase and characterized by the base sequence given as SEQ ID NO:8.

10 The terms "GPP2", "HOR2" and "YER062C" are used interchangeably and refer to a gene that encodes a cytosolic glycerol-3-phosphatase and characterized by the base sequence given as SEQ ID NO:9.

The term "GUT1" refers to a gene that encodes a cytosolic glycerol kinase and characterized by the base sequence given as SEQ ID NO:10.

15 The terms "function" or "enzyme function" refer to the catalytic activity of an enzyme in altering the energy required to perform a specific chemical reaction. It is understood that such an activity may apply to a reaction in equilibrium where the production of either product or substrate may be accomplished under suitable conditions.

20 The terms "polypeptide" and "protein" are used interchangeably.

The terms "carbon substrate" and "carbon source" refer to a carbon source capable of being metabolized by host organisms of the present invention and particularly carbon sources selected from the group consisting of monosaccharides, oligosaccharides, polysaccharides, and one-carbon substrates or mixtures thereof.

25 The terms "host cell" or "host organism" refer to a microorganism capable of receiving foreign or heterologous genes and of expressing those genes to produce an active gene product.

30 The terms "foreign gene", "foreign DNA", "heterologous gene" and "heterologous DNA" refer to genetic material native to one organism that has been placed within a host organism by various means.

The terms "recombinant organism" and "transformed host" refer to any organism having been transformed with heterologous or foreign genes. The recombinant organisms of the present invention express foreign genes encoding glycerol-3-phosphate dehydrogenase (G3PDH) and glycerol-3-phosphatase (G3P phosphatase), glycerol dehydratase (*dhaB*), and 1,3-propanediol oxidoreductase (*dhaT*) for the production of 1,3-propanediol from suitable carbon substrates.

"Gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding) and following (3' non-coding) the coding region. The terms "native" and "wild-type" refer to a gene as found in nature with its own regulatory sequences.

5 The terms "encoding" and "coding" refer to the process by which a gene, through the mechanisms of transcription and translation, produces an amino acid sequence. It is understood that the process of encoding a specific amino acid sequence includes DNA sequences that may involve base changes that do not cause a change in the encoded amino acid, or which involve base changes which
10 may alter one or more amino acids, but do not affect the functional properties of the protein encoded by the DNA sequence. It is therefore understood that the invention encompasses more than the specific exemplary sequences. Modifications to the sequence, such as deletions, insertions, or substitutions in the sequence which produce silent changes that do not substantially affect the
15 functional properties of the resulting protein molecule are also contemplated. For example, alteration in the gene sequence which reflect the degeneracy of the genetic code, or which result in the production of a chemically equivalent amino acid at a given site, are contemplated. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less
20 hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a biologically equivalent product. Nucleotide changes
25 which result in alteration of the N-terminal and C-terminal portions of the protein molecule would also not be expected to alter the activity of the protein. In some cases, it may in fact be desirable to make mutants of the sequence in order to study the effect of alteration on the biological activity of the protein. Each of the proposed modifications is well within the routine skill in the art, as is
30 determination of retention of biological activity in the encoded products. Moreover, the skilled artisan recognizes that sequences encompassed by this invention are also defined by their ability to hybridize, under stringent conditions (0.1X SSC, 0.1% SDS, 65 °C), with the sequences exemplified herein.

35 The term "expression" refers to the transcription and translation to gene product from a gene coding for the sequence of the gene product.

 The terms "plasmid", "vector", and "cassette" refer to an extra chromosomal element often carrying genes which are not part of the central metabolism of the cell, and usually in the form of circular double-stranded DNA

molecules. Such elements may be autonomously replicating sequences, genome integrating sequences, phage or nucleotide sequences, linear or circular, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is capable of introducing a promoter fragment and DNA sequence for a selected gene product along with appropriate 3' untranslated sequence into a cell. "Transformation cassette" refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that facilitate transformation of a particular host cell. "Expression cassette" refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that allow for enhanced expression of that gene in a foreign host.

The terms "transformation" and "transfection" refer to the acquisition of new genes in a cell after the incorporation of nucleic acid. The acquired genes may be integrated into chromosomal DNA or introduced as extrachromosomal replicating sequences. The term "transformant" refers to the product of a transformation.

The term "genetically altered" refers to the process of changing hereditary material by transformation or mutation.

CONSTRUCTION OF RECOMBINANT ORGANISMS:

Recombinant organisms containing the necessary genes that will encode the enzymatic pathway for the conversion of a carbon substrate to 1,3-propanediol may be constructed using techniques well known in the art. In the present invention genes encoding glycerol-3-phosphate dehydrogenase (G3PDH), glycerol-3-phosphatase (G3P phosphatase), glycerol dehydratase (*dhaB*), and 1,3-propanediol oxidoreductase (*dhaT*) were isolated from a native host such as *Klebsiella* or *Saccharomyces* and used to transform host strains such as *E. coli* DH5 α , ECL707, AA200, or W1485; the *Saccharomycetes cerevisiae* strain YPH500; or the *Klebsiella pneumoniae* strains ATCC 25955 or ECL 2106.

Isolation of Genes

Methods of obtaining desired genes from a bacterial genome are common and well known in the art of molecular biology. For example, if the sequence of the gene is known, suitable genomic libraries may be created by restriction endonuclease digestion and may be screened with probes complementary to the desired gene sequence. Once the sequence is isolated, the DNA may be amplified using standard primer directed amplification methods such as polymerase chain reaction (PCR) (U.S. 4,683,202) to obtain amounts of DNA suitable for transformation using appropriate vectors.

Alternatively, cosmid libraries may be created where large segments of genomic DNA (35-45kb) may be packaged into vectors and used to transform appropriate hosts. Cosmid vectors are unique in being able to accommodate large quantities of DNA. Generally, cosmid vectors have at least one copy of the *cos* DNA sequence which is needed for packaging and subsequent circularization of the foreign DNA. In addition to the *cos* sequence these vectors will also contain an origin of replication such as ColE1 and drug resistance markers such as a gene resistant to ampicillin or neomycin. Methods of using cosmid vectors for the transformation of suitable bacterial hosts are well described in Sambrook et al., Molecular Cloning: A Laboratory Manual, Second Edition (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).

Typically to clone cosmids, foreign DNA is isolated and ligated, using the appropriate restriction endonucleases, adjacent to the *cos* region of the cosmid vector. Cosmid vectors containing the linearized foreign DNA is then reacted with a DNA packaging vehicle such as bacteriophage λ . During the packaging process the *cos* sites are cleaved and the foreign DNA is packaged into the head portion of the bacterial viral particle. These particles are then used to transfect suitable host cells such as *E. coli*. Once injected into the cell, the foreign DNA circularizes under the influence of the *cos* sticky ends. In this manner large segments of foreign DNA can be introduced and expressed in recombinant host cells.

Isolation and cloning of genes encoding glycerol dehydratase (*dhaB*) and 1,3-propanediol oxidoreductase (*dhaT*)

Cosmid vectors and cosmid transformation methods were used within the context of the present invention to clone large segments of genomic DNA from bacterial genera known to possess genes capable of processing glycerol to 1,3-propanediol. Specifically, genomic DNA from *K. pneumoniae* ATCC 25955 was isolated by methods well known in the art and digested with the restriction enzyme Sau3A for insertion into a cosmid vector Supercos 1 and packaged using GigapackII packaging extracts. Following construction of the vector *E. coli* XL1-Blue MR cells were transformed with the cosmid DNA. Transformants were screened for the ability to convert glycerol to 1,3-propanediol by growing the cells in the presence of glycerol and analyzing the media for 1,3-propanediol formation.

Two of the 1,3-propanediol positive transformants were analyzed and the cosmids were named pKP1 and pKP2. DNA sequencing revealed extensive homology to the glycerol dehydratase gene (*dhaB*) from *C. freundii*, demonstrating that these transformants contained DNA encoding the glycerol dehydratase gene. Other 1,3-propanediol positive transformants were analyzed

and the cosmids were named pKP4 and pKP5. DNA sequencing revealed that these cosmids carried DNA encoding a diol dehydratase gene.

Although the instant invention utilizes the isolated genes from within a *Klebsiella* cosmid, alternate sources of dehydratase genes include, but are not limited to, *Citrobacter*, *Clostridia*, and *Salmonella*.

Genes encoding G3PDH and G3P phosphatase

The present invention provides genes suitable for the expression of G3PDH and G3P phosphatase activities in a host cell.

Genes encoding G3PDH are known. For example, GPD1 has been isolated from *Saccharomyces* and has the base sequence given by SEQ ID NO:5, encoding the amino acid sequence given in SEQ ID NO:11 (Wang et al., *supra*). Similarly, G3PDH activity is has also been isolated from *Saccharomyces* encoded by GPD2 having the base sequence given in SEQ ID NO:6, encoding the amino acid sequence given in SEQ ID NO:12 (Eriksson et al., *Mol. Microbiol.* 17, 95, (1995).

It is contemplated that any gene encoding a polypeptide responsible for G3PDH activity is suitable for the purposes of the present invention wherein that activity is capable of catalyzing the conversion of dihydroxyacetone phosphate (DHAP) to glycerol-3-phosphate (G3P). Further, it is contemplated that any gene encoding the amino acid sequence of G3PDH as given by any one of SEQ ID NOS:11, 12, 13, 14, 15 and 16 corresponding to the genes GPD1, GPD2, GUT2, gpsA, glpD, and the α subunit of glpABC, respectively, will be functional in the present invention wherein that amino acid sequence encompasses amino acid substitutions, deletions or additions that do not alter the function of the enzyme. It will be appreciated by the skilled person that genes encoding G3PDH isolated from other sources are also be suitable for use in the present invention. For example, genes isolated from prokaryotes include GenBank accessions M34393, M20938, L06231, U12567, L45246, L45323, L45324, L45325, U32164, and U39682; genes isolated from fungi include GenBank accessions U30625, U30876 and X56162; genes isolated from insects include GenBank accessions X61223 and X14179; and genes isolated from mammalian sources include GenBank accessions U12424, M25558 and X78593.

Genes encoding G3P phosphatase are known. For example, GPP2 has been isolated from *Saccharomyces cerevisiae* and has the base sequence given by SEQ ID NO:9 which encodes the amino acid sequence given in SEQ ID NO:17 (Norbeck et al., *J. Biol. Chem.* 271, p. 13875, 1996).

It is contemplated that any gene encoding a G3P phosphatase activity is suitable for the purposes of the present invention wherein that activity is capable

of catalyzing the conversion of glycerol-3-phosphate to glycerol. Further, it is contemplated that any gene encoding the amino acid sequence of G3P phosphatase as given by SEQ ID NOS:33 and 17 will be functional in the present invention wherein that amino acid sequence encompasses amino acid

5 substitutions, deletions or additions that do not alter the function of the enzyme. It will be appreciated by the skilled person that genes encoding G3P phosphatase isolated from other sources are also suitable for use in the present invention. For example, the dephosphorylation of glycerol-3-phosphate to yield glycerol may be achieved with one or more of the following general or specific phosphatases:

10 alkaline phosphatase (EC 3.1.3.1) [GenBank M19159, M29663, U02550 or M33965]; acid phosphatase (EC 3.1.3.2) [GenBank U51210, U19789, U28658 or L20566]; glycerol-3-phosphatase (EC 3.1.3.-) [GenBank Z38060 or U18813x11]; glucose-1-phosphatase (EC 3.1.3.10) [GenBank M33807]; glucose-6-phosphatase (EC 3.1.3.9) [GenBank U00445]; fructose-1,6-bisphosphatase (EC 3.1.3.11)

15 [GenBank X12545 or J03207] or phosphatidyl glycerol phosphate phosphatase (EC 3.1.3.27) [GenBank M23546 and M23628].

Genes encoding glycerol kinase are known. For example, GUT1 encoding the glycerol kinase from *Saccharaomyces* has been isolated and sequenced (Pavlik et al., *Curr. Genet.* 24, 21, (1993)) and the base sequence is given by SEQ ID

20 NO:10 which encodes the amino acid sequence given in SEQ ID NO:18. It will be appreciated by the skilled artisan that although glycerol kinase catalyzes the degradation of glycerol in nature the same enzyme will be able to function in the synthesis of glycerol to convert glycerol-3-phosphate to glycerol under the appropriate reaction energy conditions. Evidence exists for glycerol production

25 through a glycerol kinase. Under anaerobic or respiration-inhibited conditions, *Trypanosoma brucei* gives rise to glycerol in the presence of Glycerol-3-P and ADP. The reaction occurs in the glycosome compartment (D. Hammond, *J. Biol. Chem.* 260, 15646-15654, (1985)).

Host cells

30 Suitable host cells for the recombinant production of glycerol by the expression of G3PDH and G3P phosphatase may be either prokaryotic or eukaryotic and will be limited only by their ability to express active enzymes. Preferred hosts will be those typically useful for production of glycerol or 1,3-propanediol such as *Citrobacter*, *Enterobacter*, *Clostridium*, *Klebsiella*,

35 *Aerobacter*, *Lactobacillus*, *Aspergillus*, *Saccharomyces*, *Schizosaccharomyces*, *Zygosaccharomyces*, *Pichia*, *Kluyveromyces*, *Candida*, *Hansenula*, *Debaryomyces*, *Mucor*, *Torulopsis*, *Methylobacter*, *Escherichia*, *Salmonella*,

orientation for expression from the trc promoter of Trc99A (Pharmacia). The NcoI-BamHI fragment from pDAR1A and a second set of SpeI-RBS-NcoI linker obtained by annealing synthetic primers (SEQ ID NO:31 with SEQ ID NO:32) was inserted into the SpeI-BamHI site of pBluescript II-SK+ (Stratagene) to create pAH41. The construct pAH41 contains an ampicillin resistance gene. The NcoI-BamHI fragment from pDAR1A and a second set of SpeI-RBS-NcoI linker obtained by annealing synthetic primers (SEQ ID NO:31 with SEQ ID NO:32) was inserted into the SpeI-BamHI site of pBC-SK+ (Stratagene) to create pAH42. The construct pAH42 contains a chloramphenicol resistance gene.

10 Construction of an expression cassette for DAR1 and GPP2

An expression cassette for DAR1 and GPP2 was assembled from the individual DAR1 and GPP2 subclones described above using standard molecular biology methods. The BamHI-PstI fragment from pAH19 containing the RBS and GPP2 gene was inserted into pAH40 to create pAH43. The BamHI-PstI fragment from pAH19 containing the RBS and GPP2 gene was inserted into pAH41 to create pAH44. The same BamHI-PstI fragment from pAH19 containing the RBS and GPP2 gene was also inserted into pAH42 to create pAH45.

E. coli strain construction

20 *E. coli* W1485 is a wild-type K-12 strain (ATCC 12435). This strain was transformed with the plasmids pDT20 and pAH42 and selected on LA (Luria Agar, Difco) plates supplemented with 50 µg/mL carbencillim and 10 µg/mL chloramphenicol.

Production of 1,3-propanediol from glucose

25 *E. coli* W1485/pDT20/pAH42 was transferred from a plate to 50 mL of a medium containing per liter: 22.5 g glucose, 6.85 g K₂HPO₄, 6.3 g (NH₄)₂SO₄, 0.5 g NaHCO₃, 2.5 g NaCl, 8 g yeast extract, 8 g tryptone, 2.5 mg vitamin B₁₂, 2.5 mL modified Balch's trace-element solution, 50 mg carbencillim and 10 mg chloramphenicol, final pH 6.8 (HCl), then filter sterilized. The composition of modified Balch's trace-element solution can be found in Methods for General and
30 Molecular Bacteriology (P. Gerhardt et al., eds, p. 158, American Society for Microbiology, Washington, DC (1994)). After incubating at 37 °C, 300 rpm for 6 h, 0.5 g glucose and IPTG (final concentration = 0.2 mM) were added and shaking was reduced to 100 rpm. Samples were analyzed by GC/MS. After 24 h,
35 W1485/pDT20/pAH42 produced 1.1 g/L glycerol and 195 mg/L 1,3-propanediol.

EXAMPLE 3CLONING AND EXPRESSION OF *dhaB* AND *dhaT*IN *Saccharomyces cerevisiae*

Expression plasmids that could exist as replicating episomal elements were
 5 constructed for each of the four *dha* genes. For all expression plasmids a yeast
 ADH1 promoter was present and separated from a yeast ADH1 transcription
 terminator by fragments of DNA containing recognition sites for one or more
 restriction endonucleases. Each expression plasmid also contained the gene for
 β -lactamase for selection in *E. coli* on media containing ampicillin, an origin of
 10 replication for plasmid maintenance in *E. coli*, and a 2 micron origin of
 replication for maintenance in *S. cerevisiae*. The selectable nutritional markers
 used for yeast and present on the expression plasmids were one of the following:
 HIS3 gene encoding imidazoleglycerolphosphate dehydratase, URA3 gene
 encoding orotidine 5'-phosphate decarboxylase, TRP1 gene encoding N-(5'-
 15 phosphoribosyl)-anthranilate isomerase, and LEU2 encoding β -isopropylmalate
 dehydrogenase.

The open reading frames for *dhaT*, *dhaB3*, *dhaB2* and *dhaB1* were
 amplified from pHK28-26 (SEQ ID NO:19) by PCR using primers (SEQ ID
 NO:38 with SEQ ID NO:39, SEQ ID NO:40 with SEQ ID NO:41, SEQ ID NO:42
 20 with SEQ ID NO:43, and SEQ ID NO:44 with SEQ ID NO:45 for *dhaT*, *dhaB3*,
dhaB2 and *dhaB1*, respectively) incorporating EcoR1 sites at the 5' ends (10 mM
 Tris pH 8.3, 50 mM KCl, 1.5 mM MgCl₂, 0.0001% gelatin, 200 μ M dATP,
 200 μ M dCTP, 200 μ M dGTP, 200 μ M dTTP, 1 μ M each primer, 1-10 ng target
 DNA, 25 units/mL Amplitaq™ DNA polymerase (Perkin-Elmer Cetus, Norwalk
 25 CT)). PCR parameters were 1 min at 94 °C, 1 min at 55 °C, 1 min at 72 °C,
 35 cycles. The products were subcloned into the EcoR1 site of pHIL-D4 (Phillips
 Petroleum, Bartlesville, OK) to generate the plasmids pMP13, pMP14, pMP20
 and pMP15 containing *dhaT*, *dhaB3*, *dhaB2* and *dhaB1*, respectively.

Construction of *dhaB1* expression plasmid pMCK10

30 The 7.8 kb replicating plasmid pGADGH (Clontech, Palo Alto, CA) was
 digested with HindIII, dephosphorylated, and ligated to the *dhaB1* HindIII
 fragment from pMP15. The resulting plasmid (pMCK10) had *dhaB1* correctly
 oriented for transcription from the ADH1 promoter and contained a LEU2 marker.

Construction of *dhaB2* expression plasmid pMCK17

35 Plasmid pGADGH (Clontech, Palo Alto, CA) was digested with HindIII
 and the single-strand ends converted to EcoRI ends by ligation with HindIII-XmnI
 and EcoRI-XmnI adaptors (New England Biolabs, Beverly, MA). Selection for
 plasmids with correct EcoRI ends was achieved by ligation to a kanamycin

resistance gene on an EcoRI fragment from plasmid pUC4K (Pharmacia Biotech, Uppsala), transformation into *E. coli* strain DH5 α and selection on LB plates containing 25 μ g/mL kanamycin. The resulting plasmid (pGAD/KAN2) was digested with SnaBI and EcoRI and a 1.8 kb fragment with the ADH1 promoter was isolated. Plasmid pGBT9 (Clontech, Palo Alto, CA) was digested with SnaBI and EcoRI, and the 1.5 kb ADH1/GAL4 fragment replaced by the 1.8 kb ADH1 promoter fragment isolated from pGAD/KAN2 by digestion with SnaBI and EcoRI. The resulting vector (pMCK11) is a replicating plasmid in yeast with an ADH1 promoter and terminator and a TRP1 marker. Plasmid pMCK11 was digested with EcoRI, dephosphorylated, and ligated to the *dhaB2* EcoRI fragment from pMP20. The resulting plasmid (pMCK17) had *dhaB2* correctly oriented for transcription from the ADH1 promoter and contained a TRP1 marker.

Construction of *dhaB3* expression plasmid pMCK30

Plasmid pGBT9 (Clontech) was digested with NaeI and PvuII and the 1 kb TRP1 gene removed from this vector. The TRP1 gene was replaced by a URA3 gene donated as a 1.7 kb AatII/NaeI fragment from plasmid pRS406 (Stratagene) to give the intermediary vector pMCK32. The truncated ADH1 promoter present on pMCK32 was removed on a 1.5 kb SnaBI/EcoRI fragment, and replaced with a full-length ADH1 promoter on a 1.8 kb SnaBI/EcoRI fragment from plasmid pGAD/KAN2 to yield the vector pMCK26. The unique EcoRI site on pMCK26 was used to insert an EcoRI fragment with *dhaB3* from plasmid pMP14 to yield pMCK30. The pMCK30 replicating expression plasmid has *dhaB3* orientated for expression from the ADH1 promoter, and has a URA3 marker.

Construction of *dhaT* expression plasmid pMCK35

Plasmid pGBT9 (Clontech) was digested with NaeI and PvuII and the 1 kb TRP1 gene removed from this vector. The TRP1 gene was replaced by a HIS3 gene donated as an XmnI/NaeI fragment from plasmid pRS403 (Stratagene) to give the intermediary vector pMCK33. The truncated ADH1 promoter present on pMCK33 was removed on a 1.5 kb SnaBI/EcoRI fragment, and replaced with a full-length ADH1 promoter on a 1.8 kb SnaBI/EcoRI fragment from plasmid pGAD/KAN2 to yield the vector pMCK31. The unique EcoRI site on pMCK31 was used to insert an EcoRI fragment with *dhaT* from plasmid pMP13 to yield pMCK35. The pMCK35 replicating expression plasmid has *dhaT* orientated for expression from the ADH1 promoter, and has a HIS3 marker.

Transformation of *S. cerevisiae* with *dha* expression plasmids

S. cerevisiae strain YPH500 (*ura3-52 lys2-801 ade2-101 trp1-Δ63 his3-Δ200 leu2-Δ1*) (Sikorski R. S. and Hieter P., *Genetics* 122, 19-27, (1989)) purchased from Stratagene (La Jolla, CA) was transformed with 1-2 μ g of plasmid

DNA using a Frozen-EZ Yeast Transformation Kit (Catalog #T2001) (Zymo Research, Orange, CA). Colonies were grown on Supplemented Minimal Medium (SMM - 0.67% yeast nitrogen base without amino acids, 2% glucose) for 3-4 d at 29 °C with one or more of the following additions: adenine sulfate (20 mg/L), uracil (20 mg/L), L-tryptophan (20 mg/L), L-histidine (20 mg/L), L-leucine (30 mg/L), L-lysine (30 mg/L). Colonies were streaked on selective plates and used to inoculate liquid media.

Screening of *S. cerevisiae* transformants for *dha* genes

Chromosomal DNA from URA⁺, HIS⁺, TRP⁺, LEU⁺ transformants was analyzed by PCR using primers specific for each gene (SEQ ID NOS:38-45). The presence of all four open reading frames was confirmed.

Expression of *dhaB* and *dhaT* activity in transformed *S. cerevisiae*

The presence of active glycerol dehydratase (*dhaB*) and 1,3-propanediol oxido-reductase (*dhaT*) was demonstrated using *in vitro* enzyme assays. Additionally, western blot analysis confirmed protein expression from all four open reading frames.

Strain YPH500, transformed with the group of plasmids pMCK10, pMCK17, pMCK30 and pMCK35, was grown on Supplemented Minimal Medium containing 0.67% yeast nitrogen base without amino acids 2% glucose 20 mg/L adenine sulfate, and 30 mg/L L-lysine. Cells were homogenized and extracts assayed for *dhaB* activity. A specific activity of 0.12 units per mg protein was obtained for glycerol dehydratase, and 0.024 units per mg protein for 1,3-propanediol oxido-reductase.

EXAMPLE 4

PRODUCTION OF 1,3-PROPANEDIOL FROM D-GLUCOSE
USING RECOMBINANT *Saccharomyces cerevisiae*

S. cerevisiae YPH500, harboring the groups of plasmids pMCK10, pMCK17, pMCK30 and pMCK35, was grown in a BiostatB fermenter (B Braun Biotech, Inc.) in 1.0 L of minimal medium initially containing 20 g/L glucose, 6.7 g/L yeast nitrogen base without amino acids, 40 mg/L adenine sulfate and 60 mg/L L-lysine HCl. During the course of the growth, an additional equivalent of yeast nitrogen base, adenine and lysine was added. The fermenter was controlled at pH 5.5 with addition of 10% phosphoric acid and 2 M NaOH, 30 °C, and 40% dissolved oxygen tension through agitation control. After 38 h, the cells (OD₆₀₀ = 5.8 AU) were harvested by centrifugation and resuspended in base medium (6.7 g/L yeast nitrogen base without amino acids, 20 mg/L adenine sulfate, 30 mg/L L-lysine HCl, and 50 mM potassium phosphate buffer, pH 7.0).

Reaction mixtures containing cells ($OD_{600} = 20$ AU) in a total volume of 4 mL of base media supplemented with 0.5% glucose, 5 ug/mL coenzyme B₁₂ and 0, 10, 20, or 40 mM chloroquine were prepared, in the absence of light and oxygen (nitrogen sparging), in 10 mL crimp sealed serum bottles and incubated at 30 °C with shaking. After 30 h, aliquots were withdrawn and analyzed by HPLC. The results are shown in the Table 3.

Table 3
Production of 1,3-propanediol using recombinant *S. cerevisiae*

reaction	chloroquine (mM)	1,3-propanediol (mM)
1	0	0.2
2	10	0.2
3	20	0.3
4	40	0.7

EXAMPLE 5

10 USE OF A *S. cerevisiae* DOUBLE TRANSFORMANT FOR PRODUCTION OF 1,3-PROPANEDIOL FROM D-GLUCOSE WHERE *dhaB* AND *dhaT* ARE INTEGRATED INTO THE GENOME

Example 5 phrophetically demonstrates the transformation of *S. cerevisiae* with *dhaB1*, *dhaB2*, *dhaB3*, and *dhaT* and the stable integration of the genes into the yeast genome for the production of 1,3-propanediol from glucose.

Construction of expression cassettes

Four expression cassettes (*dhaB1*, *dhaB2*, *dhaB3*, and *dhaT*) are constructed for glucose-induced and high-level constitutive expression of these genes in yeast, *Saccharomyces cerevisiae*. These cassettes consist of: (i) the phosphoglycerate kinase (PGK) promoter from *S. cerevisiae* strain S288C; (ii) one of the genes *dhaB1*, *dhaB2*, *dhaB3*, or *dhaT*; and (iii) the PGK terminator from *S. cerevisiae* strain S288C. The PCR-based technique of gene splicing by overlap extension (Horton et al., *BioTechniques*, 8:528-535, (1990)) is used to recombine DNA sequences to generate these cassettes with seamless joints for optimal expression of each gene. These cassettes are cloned individually into a suitable vector (pLITMUS 39) with restriction sites amenable to multi-cassette cloning in yeast expression plasmids.

Construction of yeast integration vectors

Vectors used to effect the integration of expression cassettes into the yeast genome are constructed. These vectors contain the following elements: (i) a polycloning region into which expression cassettes are subcloned; (ii) a unique marker used to select for stable yeast transformants; (iii) replication origin and

selectable marker allowing gene manipulation in *E. coli* prior to transforming yeast. One integration vector contains the *URA3* auxotrophic marker (YIp352b), and a second integration vector contains the *LYS2* auxotrophic marker (pKP7).

Construction of yeast expression plasmids

- 5 Expression cassettes for *dhaB1* and *dhaB2* are subcloned into the polycloning region of the YIp352b (expression plasmid #1), and expression cassettes for *dhaB3* and *dhaT* are subcloned into the polycloning region of pKP7 (expression plasmid #2).

Transformation of yeast with expression plasmids

- 10 *S. cerevisiae* (*ura3*, *lys2*) is transformed with expression plasmid #1 using Frozen-EZ Yeast Transformation kit (Zymo Research, Orange, CA), and transformants selected on plates lacking uracil. Integration of expression cassettes for *dhaB1* and *dhaB2* is confirmed by PCR analysis of chromosomal DNA. Selected transformants are re-transformed with expression plasmid #2 using
- 15 Frozen-EZ Yeast Transformation kit, and double transformants selected on plates lacking lysine. Integration of expression cassettes for *dhaB3* and *dhaT* is confirmed by PCR analysis of chromosomal DNA. The presence of all four expression cassettes (*dhaB1*, *dhaB2*, *dhaB3*, *dhaT*) in double transformants is confirmed by PCR analysis of chromosomal DNA.

Protein production from double-transformed yeast

20 Production of proteins encoded by *dhaB1*, *dhaB2*, *dhaB3* and *dhaT* from double-transformed yeast is confirmed by Western blot analysis.

Enzyme activity from double-transformed yeast

- 25 Active glycerol dehydratase and active 1,3-propanediol dehydrogenase from double-transformed yeast is confirmed by enzyme assay as described in General Methods above.

Production of 1,3-propanediol from double-transformed yeast

Production of 1,3-propanediol from glucose in double-transformed yeast is demonstrated essentially as described in Example 4.

30 EXAMPLE 6

CONSTRUCTION OF PLASMIDS CONTAINING DARI/GPP2

OR *dhaT/dhaB1-3* AND TRANSFORMATION INTO *KLEBSIELLA* SPECIES

- 35 *K. pneumoniae* (ATCC 25955), *K. pneumoniae* (ECL2106), and *K. oxytoca* (ATCC 8724) are naturally resistant to ampicillin (up to 150 ug/mL) and kanamycin (up to 50 ug/mL), but sensitive to tetracycline (10 ug/mL) and chloramphenicol (25 ug/mL). Consequently, replicating plasmids which encode resistance to these latter two antibiotics are potentially useful as cloning vectors for these *Klebsiella* strains. The wild-type *K. pneumoniae* (ATCC 25955), the

glucose-derepressed *K. pneumonia* (ECL2106), and *K. oxytoca* (ATCC 8724) were successfully transformed to tetracycline resistance by electroporation with the moderate-copy-number plasmid, pBR322 (New England Biolabs, Beverly, MA). This was accomplished by the following procedure: Ten mL of an
5 overnight culture was inoculated into 1 L LB (1% (w/v) Bacto-tryptone (Difco, Detroit, MI), 0.5% (w/v) Bacto-yeast extract (Difco) and 0.5% (w/v) NaCl (Sigma, St. Louis, MO) and the culture was incubated at 37 °C to an OD₆₀₀ of 0.5-0.7. The cells were chilled on ice, harvested by centrifugation at 4000 x g for 15 min, and resuspended in 1 L ice-cold sterile 10% glycerol. The cells were
10 repeatedly harvested by centrifugation and progressively resuspended in 500 mL, 20 mL and, finally, 2 mL ice-cold sterile 10% glycerol. For electroporation, 40 uL of cells were mixed with 1-2 uL DNA in a chilled 0.2 cm cuvette and were pulsed at 200 Ω, 2.5 kV for 4-5 msec using a BioRad Gene Pulser (BioRad, Richmond, CA). One μL of SOC medium (2% (w/v) Bacto-tryptone (Difco),
15 0.5% (w/v) Bacto-yeast extract (Difco), 10 μM NaCl, 10 μM MgCl₂, 10 μM MgSO₄, 2.5 μM KCl and 20 μM glucose) was added to the cells and, after the suspension was transferred to a 17 x 100 mm sterile polypropylene tube, the culture was incubated for 1 hr at 37 °C, 225 rpm. Aliquots were plated on selective medium, as indicated. Analyses of the plasmid DNA from independent
20 tetracycline-resistant transformants showed the restriction endonuclease digestion patterns typical of pBR322, indicating that the vector was stably maintained after overnight culture at 37 °C in LB containing tetracycline (10 ug/mL). Thus, this vector, and derivatives such as pBR329 (ATCC 37264) which encodes resistance to ampicillin, tetracycline and chloramphenicol, may be used to introduce the
25 *DAR1/GPP2* and *dhaT/dhaB1-3* expression cassettes into *K. pneumoniae* and *K. oxytoca*.

The *DAR1* and *GPP2* genes may be obtained by PCR-mediated amplification from the *Saccharomyces cerevisiae* genome, based on their known DNA sequence. The genes are then transformed into *K. pneumoniae* or *K. oxytoca*
30 under the control of one or more promoters that may be used to direct their expression in media containing glucose. For convenience, the genes were obtained on a 2.4 kb DNA fragment obtained by digestion of plasmid pAH44 with the *PvuII* restriction endonuclease, whereby the genes are already arranged in an expression cassette under the control of the *E. coli lac* promoter. This DNA
35 fragment was ligated to *PvuII*-digested pBR329, producing the insertional inactivation of its chloramphenicol resistance gene. The ligated DNA was used to transform *E. coli* DH5α (Gibco, Gaithersburg, MD). Transformants were selected by their resistance to tetracycline (10 ug/mL) and were screened for their

sensitivity to chloramphenicol (25 ug/mL). Analysis of the plasmid DNA from tetracycline-resistant, chloramphenicol-sensitive transformants confirmed the presence of the expected plasmids, in which the P_{lac} -*dar1-gpp2* expression cassette was subcloned in either orientation into the pBR329 PvuII site. These
5 plasmids, designated pJSP1A (clockwise orientation) and pJSP1B (counter-clockwise orientation), were separately transformed by electroporation into *K. pneumonia* (ATCC 25955), *K. pneumonia* (ECL2106) and *K. oxytoca* (ATCC 8724) as described. Transformants were selected by their resistance to tetracycline (10 ug/mL) and were screened for their sensitivity to chloramphenicol
10 (25 ug/mL). Restriction analysis of the plasmids isolated from independent transformants showed only the expected digestion patterns, and confirmed that they were stably maintained at 37 °C with antibiotic selection. The expression of the *DAR1* and *GPP2* genes may be enhanced by the addition of IPTG (0.2-2.0 mM) to the growth medium.

15 The four *K. pneumoniae* *dhaB*(1-3) and *dhaT* genes may be obtained by PCR-mediated amplification from the *K. pneumoniae* genome, based on their known DNA sequence. These genes are then transformed into *K. pneumoniae* under the control of one or more promoters that may be used to direct their expression in media containing glucose. For convenience, the genes were
20 obtained on an approximately 4.0 kb DNA fragment obtained by digestion of plasmid pAH24 with the *KpnI/SacI* restriction endonucleases, whereby the genes are already arranged in an expression cassette under the control of the *E. coli lac* promoter. This DNA fragment was ligated to similarly digested pBC-KS+ (Stratagene, LaJolla, CA) and used to transform *E. coli* DH5 α . Transformants
25 were selected by their resistance to chloramphenicol (25 ug/mL) and were screened for a white colony phenotype on LB agar containing X-gal. Restriction analysis of the plasmid DNA from chloramphenicol-resistant transformants demonstrating the white colony phenotype confirmed the presence of the expected plasmid, designated pJSP2, in which the *dhaT-dhaB*(1-3) genes were subcloned
30 under the control of the *E. coli lac* promoter.

To enhance the conversion of glucose to 3G, this plasmid was separately transformed by electroporation into *K. pneumoniae* (ATCC 25955) (pJSP1A), *K. pneumoniae* (ECL2106) (pJSP1A) and *K. oxytoca* (ATCC 8724) (pJSP1A) already containing the P_{lac} -*dar1-gpp2* expression cassette. Cotransformants were
35 selected by their resistance to both tetracycline (10 ug/mL) and chloramphenicol (25 ug/mL). Restriction analysis of the plasmids isolated from independent cotransformants showed the digestion patterns expected for both pJSP1A and

pJSP2. The expression of the *DAR1*, *GPP2*, *dhaB(1-3)*, and *dhaT* genes may be enhanced by the addition of IPTG (0.2-2.0 mM) to the medium.

EXAMPLE 7

Production of 1,3 propanediol from glucose by *K. pneumoniae*

5 *Klebsiella pneumoniae* strains ECL 2106 and 2106-47, both transformed with pJSP1A, and ATCC 25955, transformed with pJSP1A and pJSP2, were grown in a 5 L Applikon fermenter under various conditions (see Table 4) for the production of 1,3-propanediol from glucose. Strain 2104-47 is a fluoroacetate-tolerant derivative of ECL 2106 which was obtained from a fluoroacetate/lactate selection plate as described in Bauer et al., *Appl. Environ. Microbiol.* 56, 1296
10 (1990). In each case, the medium used contained 50-100 mM potassium phosphate buffer, pH 7.5, 40 mM (NH₄)₂SO₄, 0.1% (w/v) yeast extract, 10 μM CoCl₂, 6.5 μM CuCl₂, 100 μM FeCl₃, 18 μM FeSO₄, 5 μM H₃BO₃, 50 μM MnCl₂, 0.1 μM Na₂MoO₄, 25 μM ZnCl₂, 0.82 mM MgSO₄, 0.9 mM CaCl₂, and 10-20 g/L
15 glucose. Additional glucose was fed, with residual glucose maintained in excess. Temperature was controlled at 37 °C and pH controlled at 7.5 with 5N KOH or NaOH. Appropriate antibiotics were included for plasmid maintenance; IPTG (isopropyl-β-D-thiogalactopyranoside) was added at the indicated concentrations as well. For anaerobic fermentations, 0.1 vvm nitrogen was sparged through the reactor; when the dO setpoint was 5%, 1 vvm air was sparged through the reactor
20 and the medium was supplemented with vitamin B12. Final concentrations and overall yields (g/g) are shown in Table 4.

Table 4

Production of 1,3 propanediol from glucose by *K. pneumoniae*

Organism	dO	IPTG, mM	vitamin B12, mg/L	Titer, g/L	Yield, g/g
25955[pJSP1A/pJSP2]	0	0.5	0	8.1	16%
25955[pJSP1A/pJSP2]	5%	0.2	0.5	5.2	4%
2106[pJSP1A]	0	0	0	4.9	17%
2106[pJSP1A]	5%	0	5	6.5	12%
2106-47[pJSP1A]	5%	0.2	0.5	10.9	12%

EXAMPLE 8Conversion of carbon substrates to 1,3-propanediol by recombinant*K. pneumoniae* containing *dar1*, *gpp2*, *dhaB*, and *dhaT*

- A. Conversion of D-fructose to 1,3-propanediol by various *K. pneumoniae* recombinant strains:
- Single colonies of *K. pneumoniae* (ATCC 25955 pJSP1A), *K. pneumoniae* (ATCC 25955 pJSP1A/pJSP2), *K. pneumoniae* (ATCC 2106 pJSP1A), and *K. pneumoniae* (ATCC 2106 pJSP1A/pJSP2) were transferred from agar plates and in separate culture tubes were subcultured overnight in Luria-Bertani (LB) broth containing the appropriate antibiotic agent(s). A 50-mL flask containing 45 mL of a steri-filtered minimal medium defined as LLMM/F which contains per liter: 10 g fructose; 1 g yeast extract; 50 mmoles potassium phosphate, pH 7.5; 40 mmoles (NH₄)₂SO₄; 0.09 mmoles calcium chloride; 2.38 mg CoCl₂•6H₂O; 0.88 mg CuCl₂•2H₂O; 27 mg FeCl₃•6H₂O; 5 mg FeSO₄•7H₂O; 0.31 mg H₃BO₃; 10 mg MnCl₂•4H₂O; 0.023 mg Na₂MoO₄•2H₂O; 3.4 mg ZnCl₂; 0.2 g MgSO₄•7H₂O. Tetracycline at 10 ug/mL was added to medium for reactions using either of the single plasmid recombinants; 10 ug/mL tetracycline and 25 ug/mL chloramphenicol for reactions using either of the double plasmid recombinants. The medium was thoroughly sparged with nitrogen prior to inoculation with 2 mL of the subculture. IPTG (I) at final concentration of 0.5 mM was added to some flasks. The flasks were capped, then incubated at 37 °C, 100 rpm in a New Brunswick Series 25 incubator/shaker. Reactions were run for at least 24 hours or until most of the carbon substrate was converted into products. Samples were analyzed by HPLC. Table 5 describes the yields of 1,3-propanediol produced from fructose by the various *Klebsiella* recombinants.

Table 5Production of 1,3-propanediol from D-fructose using recombinant *Klebsiella*

Klebsiella Strain	Medium	Conversion	[3G] (g/L)	Yield Carbon (%)
2106 pBR329	LLMM/F	100	0	0
2106 pJSP1A	LLMM/F	50	0.66	15.5
2106 pJSP1A	LLMM/F + I	100	0.11	1.4
2106 pJSP1A/pJSP2	LLMM/F	58	0.26	5
25955 pBR329	LLMM/F	100	0	0
25955 pJSP1A	LLMM/F	100	0.3	4
25955 pJSP1A	LLMM/F + I	100	0.15	2
25955 pJSP1A/pJSP2	LLMM/F	100	0.9	11
25955 pJSP1A/pJSP2	LLMM/F + I	62	1.0	20

B. Conversion of various carbon substrates to 1,3-propanediol by *K. pneumoniae* (ATCC 25955 pJSP1A/pJSP2):

- An aliquot (0.1 mL) of frozen stock cultures of *K. pneumoniae* (ATCC 25955 pJSP1A/pJSP2) was transferred to 50 mL Seed medium in a 250 mL baffled flask. The Seed medium contained per liter: 0.1 molar NaK/PO₄ buffer, pH 7.0; 3 g (NH₄)₂SO₄; 5 g glucose, 0.15 g MgSO₄·7H₂O, 10 mL 100X Trace Element solution, 25 mg chloramphenicol, 10 mg tetracycline, and 1 g yeast extract. The 100X Trace Element contained per liter: 10 g citric acid, 1.5 g CaCl₂·2H₂O, 2.8 g FeSO₄·7H₂O, 0.39 g ZnSO₄·7H₂O, 0.38 g CuSO₄·5H₂O, 0.2 g CoCl₂·6H₂O, and 0.3 g MnCl₂·4H₂O. The resulting solution was titrated to pH 7.0 with either KOH or H₂SO₄. The glucose, trace elements, antibiotics and yeast extracts were sterilized separately. The seed inoculum was grown overnight at 35 °C and 250 rpm.
- The reaction design was semi-aerobic. The system consisted of 130 mL Reaction medium in 125 mL sealed flasks that were left partially open with aluminum foil strip. The Reaction Medium contained per liter: 3 g (NH₄)₂SO₄; 20 g carbon substrate; 0.15 molar NaK/PO₄ buffer, pH 7.5; 1 g yeast extract; 0.15 g MgSO₄·7H₂O; 0.5 mmoles IPTG; 10 mL 100X Trace Element solution; 25 mg chloramphenicol; and 10 mg tetracycline. The resulting solution was titrated to pH 7.5 with KOH or H₂SO₄. The carbon sources were: D-glucose (Glc); D-fructose (Frc); D-lactose (Lac); D-sucrose (Suc); D-maltose (Mal); and D-mannitol (Man). A few glass beads were included in the medium to improve mixing. The reactions were initiated by addition of seed inoculum so that the optical density of the cell suspension started at 0.1 AU as measured at λ₆₀₀ nm. The flasks were incubated at 35 °C: 250 rpm. 3G production was measured by HPLC after 24 hr. Table 6 describes the yields of 1,3-propanediol produced from the various carbon substrates.

30

Table 6
Production of 1,3-propanediol from various carbon substrates
using recombinant *Klebsiella* 25955 pJSP1A/pJSP2

Carbon Substrate	1,3-Propanediol (g/L)		
	Expt. 1	Expt. 2	Expt 3
Glc	0.89	1	1.6
Frc	0.19	0.23	0.24
Lac	0.15	0.58	0.56
Suc	0.88	0.62	
Mal	0.05	0.03	0.02
Man	0.03	0.05	0.04

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

(A) ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
(B) STREET: 1007 MARKET STREET
(C) CITY: WILMINGTON
(D) STATE: DELAWARE
(E) COUNTRY: U.S.A.
(F) ZIP: 19898
(G) TELEPHONE: 302-892-8112
(H) TELEFAX: 302-773-0164
(I) TELEX: 6717325

(A) ADDRESSEE: GENENCOR INTERNATIONAL, INC.
(B) STREET: 4 CAMBRIDGE PLACE
1870 SOUTH WINTON ROAD
(C) CITY: ROCHESTER
(D) STATE: NEW YORK
(E) COUNTRY: U.S.A.
(F) POSTAL CODE (ZIP): 14618

(ii) TITLE OF INVENTION: METHOD FOR THE RECOMBINANT
PRODUCTION OF 1,3-PROPANEDIOL

(iii) NUMBER OF SEQUENCES: 49

(iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: 3.50 INCH DISKETTE
(B) COMPUTER: IBM PC COMPATIBLE
(C) OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
(D) SOFTWARE: MICROSOFT WORD VERSION 7.0A

(v) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vi) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: 60/030,601
(B) FILING DATE: NOVEMBER 13, 1996

(vii) ATTORNEY/AGENT INFORMATION:

(A) NAME: FLOYD, LINDA AXAMETHY
(B) REGISTRATION NO.: 33,692
(C) REFERENCE/DOCKET NUMBER: CR-9982

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1668 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: DHAB1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATGAAAAGAT CAAACGATT TGCAGTACTG GCCCAGCGCC CCGTCAATCA GGACGGGCTG	60
ATTGGCGAGT GGCCTGAAGA GGGGCTGATC GCCATGGACA GCCCCTTTGA CCCGGTCTCT	120
TCAGTAAAG TGGACAACGG TCTGATCGTC GAACTGGACG GCAAACGCCG GGACCAGTTT	180
GACATGATCG ACCGATTAT CGCCGATTAC GCGATCAACG TTGAGCGCAC AGAGCAGGCA	240
ATGCGCCTGG AGGCGGTGGA AATAGCCCGT ATGCTGGTGG ATATTCACGT CAGCCGGGAG	300
GAGATCATTG CCATCACTAC CGCCATCACG CCGGCCAAAG CGGTCGAGGT GATGGCGCAG	360
ATGAACGTGG TGGAGATGAT GATGGCGCTG CAGAAGATGC GTGCCC GCCG GACCCCTCC	420
AACCAGTGCC ACGTCACCAA TCTCAAAGAT AATCCGGTGC AGATTGCCGC TGACGCCGCC	480
GAGGCCGGGA TCCGCGGCTT CTCAGAACAG GAGACCACGG TCGGTATCGC GCGCTACGCG	540
CCGTTTAACG CCCTGGCGCT GTTGGTCGGT TCGCAGTGC GCGCCCCCG CGTGTTGACG	600
CAGTGCTCGG TGGAAGAGGC CACCGAGCTG GAGCTGGGCA TCGGTGGCTT AACCAGCTAC	660
GCCGAGACGG TGTCGGTCTA CGGCACCGAA GCGGTATTTA CCGACGGCGA TGATACGCCG	720
TGGTCAAAGG CGTTCCTCGC CTCGGCCTAC GCCTCCCGCG GGTGAAAAT GCGCTACACC	780
TCCGGCACCG GATCCGAAGC GCTGATGGGC TATTCGGAGA GCAAGTCGAT GCTCTACCTC	840
GAATCGCGCT GCATCTTCAT TACTAAAGGC GCCGGGGTTC AGGGACTGCA AAACGGCGCG	900
GTGAGCTGTA TCGGCATGAC CGGCGCTGTG CCGTCGGGCA TTCGGGCGGT GCTGGCGGAA	960
AACCTGATCG CCTCTATGCT CGACCTCGAA GTGGCGTCCG CCAACGACCA GACTTTCTCC	1020
CACTCGGATA TTCGCCGCAC CGCGCGCACC CTGATGCAGA TGCTGCCGGG CACCGACTTT	1080
ATTTTCTCCG GCTACAGCGC GGTGCCGAAC TACGACAACA TGTTCCGGG CTCGAACTTC	1140
GATGCGGAAG ATTTTGATGA TTACAACATC CTGCAGCGTG ACCTGATGGT TGACGGCGGC	1200
CTGCGTCCGG TGACCGAGGC GGAAACCATT GCCATTCGCC AGAAAGCGGC GCGGGCGATC	1260
CAGGCGGTTT TCCGCGAGCT GGGGCTGCCG CCAATCGCCG ACGAGGAGGT GGAGCCGCC	1320
ACCTACGCGC ACGGCAGCAA CGAGATGCCG CCGCGTAACG TGGTGGAGGA TCTGAGTGCG	1380
GTGGAAGAGA TGATGAAGCG CAACATCACC GGCCTCGATA TTGTCGGCGC GCTGAGCCGC	1440

AGCGGCTTTG AGGATATCGC CAGCAATATT CTCAATATGC TGC GCCAGCG GGTCACCGGC 1500
 GATTACCTGC AGACCTCGGC CATTCTCGAT CGGCAGTTCG AGGTGGTGAG TCGGGTCAAC 1560
 GACATCAATG ACTATCAGGG GCCGGGCACC GGCTATCGCA TCTCTGCCGA ACGCTGGGCG 1620
 GAGATCAAAA ATATTCCGGG CGTGGTTCAG CCCGACACCA TTGAATAA 1668

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 585 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: DHAB2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GTGCAACAGA CAACCCAAAT TCAGCCCTCT TTTACCCTGA AAACCCGCGA GGGCGGGGTA 60
 GCTTCTGCCG ATGAACGCGC CGATGAAGTG GTGATCGGCG TCGGCCCTGC CTTGATAAA 120
 CACCAGCATC AACTCTGAT CGATATGCCC CATGGCGCGA TCCTCAAAGA GCTGATTGCC 180
 GGGGTGGAAG AAGAGGGGCT TCACGCCCGG GTGGTGCGCA TTCTGCGCAC GTCCGACGTC 240
 TCCTTTATGG CCTGGGATGC GGCCAACCTG AGCGGCTCGG GGATCGGCAT CGGTATCCAG 300
 TCGAAGGGGA CCACGGTCAT CCATCAGCGC GATCTGCTGC CGCTCAGCAA CCTGGAGCTG 360
 TTCTCCAGG CGCCGCTGCT GACGCTGGAG ACCTACCGGC AGATTGGCAA AAACGCTGCG 420
 CGCTATGCGC GCAAAGAGTC ACCTTCGCCG GTGCCGGTGG TGAACGATCA GATGGTGCGG 480
 CCGAAATTTA TGGCAAAGC CGCGCTATTT CATATCAAAG AGACCAAACA TGTGGTGCAG 540
 GACGCCGAGC CCGTCACCCT GCACATCGAC TTAGTAAGGG AGTGA 585

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 426 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: DHAB3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATGAGCGAGA AAACCATGCG CGTGCAGGAT TATCCGTTAG CCACCCGCTG CCCGGAGCAT 60
 ATCCTGACGC CTACCGGCAA ACCATTGACC GATATTACCC TCGAGAAGGT GCTCTCTGGC 120
 GAGGTGGGCC CGCAGGATGT GCGGATCTCC CGCCAGACCC TTGAGTACCA GGCGCAGATT 180
 GCCGAGCAGA TGCAGCGCCA TCGGGTGGCG CGCAATTTCC GCCGCGCGGC GGAGCTTATC 240

GCCATTCCTG ACGAGCGCAT TCTGGCTATC TATAACGCGC TCGCCCCGTT CCGCTCCTCG 300
 CAGGCGGAGC TGCTGGCGAT CGCCGACGAG CTGGAGCACA CCTGGCATGC GACAGTGAAT 360
 GCCGCCTTTG TCCGGGAGTC GCGGAAGTG TATCAGCAGC GGCATAAGCT GCGTAAAGGA 420
 AGCTAA 426

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1164 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: DHAT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

ATGAGCTATC GTATGTTTGA TTATCTGGTG CCAAACGTTA ACTTTTTTGG CCCCAACGCC 60
 ATTTCCGTAG TCGGCGAACG CTGCCAGCTG CTGGGGGGGA AAAAAGCCCT GCTGGTCACC 120
 GACAAAGGCC TCGGGGCAAT TAAAGATGGC GCGGTGGACA AAACCCTGCA TTATCTGCGG 180
 GAGGCCGGGA TCGAGGTGGC GATCTTTGAC GCGCTCGAGC CGAACCCGAA AGACACCAAC 240
 GTGCGCGACG GCCTCGCCGT GTTTCGCCGC GAACAGTGGC ACATCATCGT CACCGTGGGC 300
 GCGGCAGGCC CGCACGATTG CGGCAAAGGC ATCGGCATCG CCGCCACCCA TGAGGGCGAT 360
 CTGTACCAGT ATGCCGGAAT CGAGACCCTG ACCAACCCGC TGCCGCCTAT CGTCGCGGTC 420
 AATACCACCG CCGGCACCGC CAGCGAGGTC ACCCGCCACT GCGTCCTGAC CAACACCGAA 480
 ACCAAAGTGA AGTTTGTGAT CGTCAGCTGG CGCAAAGTGC CGTCGGTCTC TATCAACGAT 540
 CCACTGCTGA TGATCGGTAA ACCGGCCGCC CTGACCGCGG CGACCGGGAT GGATGCCCTG 600
 ACCCAGCCG TAGAGGCCA TATCTCCAAA GACGCTAACC CGGTGACGGA CGCCGCCGCC 660
 ATGCAGGCGA TCCGCCTCAT CGCCGCAAC CTGCGCCAGG CCGTGGCCCT CGGCAGCAAT 720
 CTGCAGGCGC GGGAAAACAT GGCCTATGCT TCTCTGCTGG CCGGGATGGC TTTCAATAAC 780
 GCCAACCTCG GCTACGTGCA CGCCATGGCG CACCAGCTGG GCGGCCTGTA CGACATGCCG 840
 CACGGCGTGG CCAACGCTGT CCTGCTGCCG CATGTGGCGC GCTACAACCT GATCGCCAAC 900
 CCGGAGAAAT TCGCCGATAT CGCTGAACTG ATGGGCGAAA ATATCACCGG ACTGTCCACT 960
 CTCGACGCGG CGGAAAAAGC CATCGCCGCT ATCAGCGGTC TGTCGATGGA TATCGGTATT 1020
 CCGCAGCATC TGCGCGATCT GGGGGTAAAA GAGGCCGACT TCCCCTACAT GGCGGAGATG 1080
 GCTCTAAAAG ACGGCAATGC GTTCTCGAAC CCGCGTAAAG GCAACGAGCA GGAGATTGCC 1140
 GCGATTTTCC GCCAGGCATT CTGA 1164

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1380 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: GPD1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```

CTTTAATTTT CTTTATCTT ACTCTCCTAC ATAAGACATC AAGAAACAAT TGTATATTGT   60
ACACCCCCCC CCTCCACAAA CACAAATATT GATAATATAA AGATGTCTGC TGCTGCTGAT  120
AGATTAAACT TAACTTCCGG CCACTTGAAT GCTGGTAGAA AGAGAAGTTC CTCTTCTGTT  180
TCTTTGAAGG CTGCCGAAAA GCCTTTCAAG GTTACTGTGA TTGGATCTGG TAACTGGGGT  240
ACTACTATTG CCAAGGTGGT TGCCGAAAAT TGTAAGGGAT ACCCAGAAGT TTTGCTCCA  300
ATAGTACAAA TGTGGGTGTT CGAAGAAGAG ATCAATGGTG AAAAATTGAC TGAAATCATA  360
AATACTAGAC ATCAAAACGT GAAATACTTG CCTGGCATCA CTCTACCCGA CAATTGGTT  420
GCTAATCCAG ACTTGATTGA TTCAGTCAAG GATGTCGACA TCATCGTTTT CAACATTCCA  480
CATCAATTTT TGCCCCGTAT CTGTAGCCAA TTGAAAGGTC ATGTTGATTC ACACGTCAGA  540
GCTATCTCCT GTCTAAAGGG TTTTGAAGTT GGTGCTAAAG GTGTCCAATT GCTATCTCT  600
TACATCACTG AGGAACTAGG TATTCAATGT GGTGCTCTAT CTGGTGCTAA CATTGCCACC  660
GAAGTCGCTC AAGAACACTG GTCTGAAACA ACAGTTGCTT ACCACATTCC AAAGGATTT  720
AGAGGCGAGG GCAAGGACGT CGACCATAAG GTTCTAAAGG CCTTGTTCCA CAGACCTTAC  780
TTCCACGTTA GTGTCATCGA AGATGTTGCT GGTATCTCCA TCTGTGGTGC TTTGAAGAAC  840
GTTGTTGCCT TAGGTTGTGG TTTCGTCGAA GGTCTAGGCT GGGGTAACAA CGCTTCTGCT  900
GCCATCCAAA GAGTCGGTTT GGGTGAGATC ATCAGATTCG GTCAAATGTT TTTCCAGAA  960
TCTAGAGAAG AAACATACTA CCAAGAGTCT GCTGGTGTG CTGATTTGAT CACCACCTGC 1020
GCTGGTGGTA GAAACGTCAA GGTGCTAGG CTAATGGCTA CTTCTGGTAA GGACGCCTGG 1080
GAATGTGAAA AGGAGTTGTT GAATGGCCAA TCCGCTCAAG GTTTAATTAC CTGCAAAGAA 1140
GTTACGAAT GGTGGAAC ATGTGGCTCT GTCGAAGACT TCCCATTATT TGAAGCCGTA 1200
TACCAAATCG TTTACAACAA CTACCCAATG AAGAACCTGC CGGACATGAT TGAAGAATTA 1260
GATCTACATG AAGATTAGAT TTATTGGAGA AAGATAACAT ATCATACTTC CCCCACTTTT 1320
TTCGAGGCTC TTCTATATCA TATTCATAAA TTAGCATTAT GTCATTTCTC ATAACTACTT 1380

```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2946 base pairs
- (B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: GPD2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GAATTCGAGC CTGAAGTGCT GATTACCTTC AGGTAGACTT CATCTTGACC CATCAACCCC	60
AGCGTCAATC CTGCAAATAC ACCACCCAGC AGCACTAGGA TGATAGAGAT AATATAGTAC	120
GTGGTAACGC TTGCCTCATC ACCTACGCTA TGGCCGGAAT CGGCAACATC CCTAGAATTG	180
AGTAGCTGTG ATCCGGATAA CAACGGCAGT GAATATATCT TCGGTATCGT AAAGATGTGA	240
TATAAGATGA TGTATACCCA ATGAGGAGCG CCTGATCGTG ACCTAGACCT TAGTGGCAAA	300
AACGACATAT CTATTATAGT GGGGAGAGTT TCGTGCAAAT AACAGACGCA GCAGCAAGTA	360
ACTGTGACGA TATCAACTCT TTTTTTATTA TGTAATAAGC AAACAAGCAC GAATGGGGAA	420
AGCCTATGTG CAATCACCAA GGTCGTCCCT TTTTTCCCAT TTGCTAATTT AGAATTTAAA	480
GAAACCAAAA GAATGAAGAA AGAAAACAAA TACTAGCCCT AACCTGACT TCGTTTCTAT	540
GATAATACCC TGCTTTAATG AACGGTATGC CCTAGGGTAT ATCTCACTCT GTACGTTACA	600
AACTCCGGTT ATTTTATCGG AACATCCGAG CACCCGCGCC TTCCTCAACC CAGGCACCGC	660
CCCAGGTAAC CGTGCGCGAT GAGCTAATCC TGAGCCATCA CCCACCCAC CCGTTGATGA	720
CAGCAATTCG GGAGGGCGAA AATAAACTG GAGCAAGGAA TTACCATCAC CGTCACCATC	780
ACCATCATAT CGCCTTAGCC TCTAGCCATA GCCATCATGC AAGCGTGAT CTTCTAAGAT	840
TCAGTCATCA TCATTACCGA GTTTGTTTTT CTTACATGA TGAAGAAGGT TTGAGTATGC	900
TCGAAACAAT AAGACGACGA TGGCTCTGCC ATTGGTTATA TTACGCTTTT GCGGCGAGGT	960
GCCGATGGGT TGCTGAGGGG AAGAGTGTTT AGCTTACGGA CCTATTGCCA TTGTTATTCC	1020
GATTAATCTA TTGTTACGCA GCTCTTCTCT ACCCTGTCAT TCTAGTATTT TTTTTTTTTT	1080
TTTTTGTTT TACTTTTTTT TCTTCTTGCC TTTTTTCTT GTTACTTTTT TTCTAGTTTT	1140
TTTTTCCTCC ACTAAGCTTT TTCCTTGATT TATCCTTGGG TTCTTCTTTC TACTCCTTTA	1200
GATTTTTTTT TTATATATTA ATTTTAAAGT TTATGTATTT TGGTAGATTC AATTCTCTTT	1260
CCCTTTCCTT TTCCTTCGCT CCCCTTCCTT ATCAATGCTT GCTGTCAGAA GATTAACAAG	1320
ATACACATTC CTTAAGCGAA CGCATCCGGT GTTATATACT CGTCGTGCAT ATAAAATTTT	1380
GCCTTCAAGA TCTACTTTCC TAAGAAGATC ATTATTACAA ACACAACGTC ACTCAAAGAT	1440
GACTGCTCAT ACTAATATCA AACAGCACAA AACTGTGCAT GAGGACCATC CTATCAGAAG	1500
ATCGGACTCT GCCGTGTCAA TTGTACATTT GAAACGTGCG CCCTTCAAGG TTACAGTGAT	1560
TGGTTCTGGT AACTGGGGGA CCACCATCGC CAAAGTCATT GCGGAAAACA CAGAATTGCA	1620
TTCCCATATC TTCGAGCCAG AGGTGAGAAT GTGGGTTTTT GATGAAAAGA TCGGCGACGA	1680

AAATCTGACG GATATCATAA ATACAAGACA CCAGAACGTT AAATATCTAC CCAATATTGA 1740
 CCTGCCCCAT AATCTAGTGG CCGATCCTGA TCTTTTACAC TCCATCAAGG GTGCTGACAT 1800
 CCTTGTTTTT AACATCCCTC ATCAATTTTT ACCAAACATA GTCAAACAAT TGCAAGGCCA 1860
 CGTGGCCCCCT CATGTAAGGG CCATCTCGTG TCTAAAAGGG TTCGAGTTGG GCTCCAAGGG 1920
 TGTGCAATTG CTATCCTCCT ATGTTACTGA TGAGTTAGGA ATCCAATGTG GCGCACTATC 1980
 TGGTGCAAAC TTGGCACCAG AAGTGGCCAA GGAGCATTGG TCCGAAACCA CCGTGGCTTA 2040
 CCAACTACCA AAGGATTATC AAGGTGATGG CAAGGATGTA GATCATAAGA TTTTGAAATT 2100
 GCTGTTCCAC AGACCTTACT TCCACGTCAA TGTCATCGAT GATGTTGCTG GTATATCCAT 2160
 TGCCGGTGCC TTGAAGAACG TCGTGGCACT TGCATGTGGT TTCGTAGAAG GTATGGGATG 2220
 GGGTAACAAT GCCTCCGCAG CCATTCAAAG GCTGGGTTTA GGTGAAATTA TCAAGTTCGG 2280
 TAGAATGTTT TTCCCAGAAT CCAAAGTCGA GACCTACTAT CAAGAATCCG CTGGTGTTGC 2340
 AGATCTGATC ACCACCTGCT CAGGCGGTAG AAACGTCAAG GTTGCCACAT ACATGGCCAA 2400
 GACCGGTAAG TCAGCCTTGG AAGCAGAAAA GGAATTGCTT AACGGTCAAT CCGCCCAAGG 2460
 GATAATCACA TGCAGAGAAG TTCACGAGTG GCTACAAACA TGTGAGTTGA CCCAAGAATT 2520
 CCCAATTATT CGAGGCAGTC TACCAGATAG TCTACAACAA CGTCCGCATG GAAGACCTAC 2580
 CGGAGATGAT TGAAGAGCTA GACATCGATG ACGAATAGAC ACTCTCCCCC CCCCTCCCCC 2640
 TCTGATCTTT CCTGTTGCCT CTTTTTCCCC CAACCAATTT ATCATTATAC ACAAGTTCTA 2700
 CAACTACTAC TAGTAACATT ACTACAGTTA TTATAATTTT CTATTCTCTT TTTCTTTAAG 2760
 AATCTATCAT TAACGTTAAT TTCTATATAT ACATAACTAC CATTATACAC GCTATTATCG 2820
 TTTACATATC ACATCACCGT TAATGAAAGA TACGACACCC TGTACACTAA CACAATTAAA 2880
 TAATCGCCAT AACCTTTTCT GTTATCTATA GCCCTTAAAG CTGTTTCTTC GAGCTTTTCA 2940
 CTGCAG 2946

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3178 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: GUT2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CTGCAGAACT TCGTCTGCTC TGTGCCCATC CTCGCGGTTA GAAAGAAGCT GAATTGTTTC 60
 ATGCGCAAGG GCATCAGCGA GTGACCAATA ATCACTGCAC TAATTCCTTT TTAGCAACAC 120
 ATACTTATAT ACAGCACCAG ACCTTATGTC TTTTCTCTGC TCCGATACGT TATCCCACCC 180
 AACTTTTATT TCAGTTTGG CAGGGGAAAT TTCACAACCC CGCAGCTAA AAATCGTATT 240

TAAACTTAAA AGAGAACAGC CACAAATAGG GAACTTTGGT CTAAACGAAG GACTCTCCCT 300
CCCTTATCTT GACCGTGCTA TTGCCATCAC TGCTACAAGA CTAAATACGT ACTAATATAT 360
GTTTTCGGTA ACGAGAAGAA GAGCTGCCGG TGCAGCTGCT GCCATGGCCA CAGCCACGGG 420
GACGCTGTAC TGGATGACTA GCCAAGGTGA TAGGCCGTTA GTGCACAATG ACCCGAGCTA 480
CATGGTGCAA TTCCCCACCG CCGCTCCACC GGCAGGTCTC TAGACGAGAC CTGCTGGACC 540
GTCTGGACAA GACGCATCAA TTCGACGTGT TGATCATCGG TGGCGGGGCC ACGGGGACAG 600
GATGTGCCCT AGATGCTGCG ACCAGGGGAC TCAATGTGGC CCTTGTGAA AAGGGGGATT 660
TTGCCTCGGG AACGTCGTCC AAATCTACCA AGATGATTCA CGGTGGGGTG CGGTACTTAG 720
AGAAGGCCCTT CTGGGAGTTC TCCAAGGCAC AACTGGATCT GGTTCATCGAG GCACTCAACG 780
AGCGTAAACA TCTTATCAAC ACTGCCCTC ACCTGTGCAC GGTGCTACCA ATTCTGATCC 840
CCATCTACAG CACCTGGCAG GTCCCGTACA TCTATATGGG CTGTAAATTC TACGATTCT 900
TTGGCGGTTT CCAAACTTG AAAAAATCAT ACCTACTGTC CAAATCCGCC ACCGTGGAGA 960
AGGCTCCCAT GCTTACCACA GACAATTTAA AGGCCTCGCT TGTGTACCAT GATGGGTCTT 1020
TTAAGACTC GCGTTTGAAC GCCACTTTAG CCATCACGGG TGTGGAGAAC GGCCTACCG 1080
TCTTGATCTA TGTCGAGGTA CAAAAATTGA TCAAAGACCC AACTTCTGGT AAGGTTATCG 1140
GTGCCGAGGC CCGGGACGTT GAGACTAATG AGCTTGTCAG AATCAACGCT AAATGTGTGG 1200
TCAATGCCAC GGGCCCATAC AGTGACGCCA TTTTGCAAAT GGACCGCAAC CCATCCGGTC 1260
TGCCGACTC CCCGCTAAAC GACAACTCCA AGATCAAGTC GACTTTCAAT CAAATCTCCG 1320
TCATGGACCC GAAAATGGTC ATCCCATCTA TTGGCGTTCA CATCGTATTG CCCTCTTTTT 1380
ACTCCCCGAA GGATATGGGT TTGTTGGACG TCAGAACCTC TGATGGCAGA GTGATGTTCT 1440
TTTTACCTTG GCAGGGCAAA GTCCTTGCCG GCACCACAGA CATCCCACTA AAGCAAGTCC 1500
CAGAAAACCC TATGCCTACA GAGGCTGATA TTCAAGATAT CTTGAAAGAA CTACAGCACT 1560
ATATCGAATT CCCCGTGAAG AGAGAAGACG TGCTAAGTGC ATGGGCTGGT GTCAGACCTT 1620
TGGTCAGAGA TCCACGTACA ATCCCCGAG ACGGGAAGAA GGGCTCTGCC ACTCAGGGCG 1680
TGGTAAGATC CCACTTCTTG TCACTTCGG ATAATGGCCT AATTACTATT GCAGGTGGTA 1740
AATGGACTAC TTACAGACAA ATGGCTGAGG AAACAGTCGA CAAAGTTGTC GAAGTTGGCG 1800
GATTCCACAA CCTGAAACCT TGTCACACAA GAGATATTAA GCTTGCTGGT GCAGAAGAAT 1860
GGACGCAAAA CTATGTGGCT TTATTGGCTC AAAACTACCA TTTATCATCA AAAATGTCCA 1920
ACTACTTGGT TCAAACTAC GGAACCCGTT CCTCTATCAT TTGCGAATTT TTCAAAGAAT 1980
CCATGGAAAA TAAACTGCCT TTGTCCTTAG CCGACAAGGA AAATAACGTA ATCTACTCTA 2040
GCGAGGAGAA CAACTTGGTC AATTTTGATA CTTTCAGATA TCCATTCACA ATCGGTGAGT 2100
TAAAGTATTC CATGCAGTAC GAATATTGTA GAACTCCCTT GGAATTCCTT TTAAGAAGAA 2160
CAAGATTCGC CTTCTTGAC GCCAAGGAAG CTTTGAATGC CGTGCATGCC ACCGTCAAAG 2220

```

TTATGGGTGA TGAGTTCAAT TGGTCGGAGA AAAAGAGGCA GTGGGAAC TT GAAAAAACTG 2280
TGAAC TTCAT CCAAGGACGT TTCGGTGTCT AAATCGATCA TGATAGTTAA GGGTGACAAA 2340
GATAACATTC ACAAGAGTAA TAATAATGGT AATGATGATA ATAATAATAA TGATAGTAAT 2400
AACATAATA ATAATGGTGG TAATGGCAAT GAAATCGCTA TTATTACCTA TTTTCCTTAA 2460
TGGAAGAGTT AAAGTAACT AAAAAAACTA CAAAAATATA TGAAGAAAAA AAAAAAAGA 2520
GGTAATAGAC TCTACTACTA CAATTGATCT TCAAATTATG ACCTTCCTAG TGTTTATATT 2580
CTATTTCCAA TACATAATAT AATCTATATA ATCATTGCTG GTAGACTTCC GTTTTAATAT 2640
CGTTTTAATT ATCCCTTTA TCTCTAGTCT AGTTTATCA TAAATATAG AAACACTAAA 2700
TAATATTCTT CAAACGGTCC TGGTGCATAC GCAATACATA TTTATGGTGC AAAAAAAAAA 2760
ATGGA AAAT TTGCTAGTCA TAAACCCTTT CATAAAACAA TACGTAGACA TCGCTACTTG 2820
AAATTTTCAA GTTTTTATCA GATCCATGTT TCCTATCTGC CTGACAACC TCATCGTCGA 2880
AATAGTACCA TTTAGAACGC CCAATATTCA CATGTGTTC AAGGTCTTTA TTCACCAGTG 2940
ACGTGTAATG GCCATGATTA ATGTGCCTGT ATGGTTAACC ACTCCAAATA GCTTATATTT 3000
CATAGTGTCA TTGTTTTTCA ATATAATGTT TAGTATCAAT GGATATGTTA CGACGGTGTT 3060
ATTTTCTTG GTCAAATCGT AATAAAATCT CGATAAATGG ATGACTAAGA TTTTGGTAA 3120
AGTTACAAA TTTATCGTTT TCACTGTTGT CAATTTTTTG TTCTTGTAAT CACTCGAG 3178

```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 816 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: GPPI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```

ATGAAACGTT TCAATGTTTT AAAATATATC AGAACACAA AAGCAAATAT ACAAACCATC 60
GCAATGCCTT TGACCACAAA ACCTTTATCT TTGAAATCA ACGCCGCTCT ATTCGATGTT 120
GACGGTACCA TCATCATCTC TCAACCAGCC ATTGCTGCTT TCTGGAGAGA TTTCGGTAAA 180
GACAAGCCTT ACTTCGATGC CGAACACGTT ATTCACATCT CTCACGGTTG GAGAACTTAC 240
GATGCCATTG CCAAGTTCGC TCCAGACTTT GCTGATGAAG AATACGTAA CAAGCTAGAA 300
GGTGAAATCC CAGAAAAGTA CGGTGAACAC TCCATCGAAG TTCCAGGTGC TGCAAGTTG 360
TGTAATGCTT TGAACGCCTT GCCAAAGGAA AAATGGGCTG TCGCCACCTC TGGTACCCGT 420
GACATGGCCA AGAAATGGTT CGACATTTTG AAGATCAAGA GACCAGAATA CTTCATCACC 480
GCCAATGATG TCAAGCAAGG TAAGCCTCAC CCAGAACCAT ACTTAAAGGG TAGAAACGGT 540

```

TTGGGTTTCC CAATTAATGA ACAAGACCCA TCCAAATCTA AGGTTGTTGT CTTTGAAGAC 600
 GCACCAGCTG GTATTGCTGC TGGTAAGGCT GCTGGCTGTA AAATCGTTGG TATTGCTACC 660
 ACTTTCGATT TGGACTTCTT GAAGGAAAAG GGTGTGACA TCATTGTCAA GAACCACGAA 720
 TCTATCAGAG TCGGTGAATA CAACGCTGAA ACCGATGAAG TCGAATTGAT CTTTGATGAC 780
 TACTTATACG CTAAGGATGA CTTGTTGAAA TGGTAA 816

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 753 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: GPP2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ATGGGATTGA CTAATAAACC TCTATCTTTC AAAGTTAAGC CCGCTTTGTT CGACGTCGAC 60
 GGTACCATTA TCATCTCTCA ACCAGCCATT GCTGCATTCT GGAGGGATTT CGGTAAGGAC 120
 AAACCTTATT TCGATGCTGA ACACGTTATC CAAGTCTCGC ATGGTTGGAG AACGTTTGAT 180
 GCCATTGCTA AGTTCGCTCC AGACTTTGCC AATGAAGAGT ATGTTAACAA ATTAGAAGCT 240
 GAAATTCGG TCAAGTACGG TGAAAAATCC ATTGAAGTCC CAGGTGCAGT TAAGCTGTGC 300
 AACGCTTTGA ACGCTCTACC AAAAGAGAAA TGGGCTGTGG CAACTTCGG TACCCGTGAT 360
 ATGGCACAAA AATGGTTCGA GCATCTGGGA ATCAGGAGAC CAAAGTACTT CATTACCGCT 420
 AATGATGTCA AACAGGGTAA GCCTCATCCA GAACCATATC TGAAGGGCAG GAATGGCTTA 480
 GGATATCCGA TCAATGAGCA AGACCCTTCC AAATCTAAGG TAGTAGTATT TGAAGACGCT 540
 CCAGCAGGTA TTGCCGCCGG AAAAGCCGCC GGTGTAAGA TCATTGGTAT TGCCACTACT 600
 TTCGACTTGG ACTTCCTAAA GGAAAAAGGC TGTGACATCA TTGTCAAAAA CCACGAATCC 660
 ATCAGAGTTG GCGGCTACAA TGCCGAAACA GACGAAGTTG AATTCATTTT TGACGACTAC 720
 TTATATGCTA AGGACGATCT GTTGAAATGG TAA 753

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2520 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: GUT1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TGTATTGGCC ACGATAACCA CCCTTTGTAT ACTGTTTTTG TTTTTCACAT GGTAATAAC	60
GACTTTTATT AAACAACGTA TGTA AAAACA TAACAAGAAT CTACCCATAC AGGCCATTTC	120
GTAATTCTTC TCTTCTAATT GGAGTAAAC CATCAATTAA AGGGTGTGGA GTAGCATAGT	180
GAGGGGCTGA CTGCATTGAC AAAAAAATTG AAAAAAAAAA AGGAAAAGGA AAGGAAAAAA	240
AGACAGCCAA GACTTTTAGA ACGGATAAGG TGTAATAAAA TGTGGGGGGA TGCCTGTTCT	300
CGAACCATAT AAAATATACC ATGTGGTTTG AGTTGTGGCC GGAACATATC AAATAGTTAT	360
ATGTTTCCCT CTCTCTCCG ACTTGTAGTA TTCTCCAAAC GTTACATATT CCGATCAAGC	420
CAGCGCCTTT AACTAGTTT AAAACAAGAA CAGAGCCGTA TGTCCAAAT AATGGAAGAT	480
TTACGAAGTG ACTACGTCCC GCTTATCGCC AGTATTGATG TAGGAACGAC CTCATCCAGA	540
TGCATTCTGT TCAACAGATG GGGCCAGGAC GTTTCAAAAC ACCAAATTGA ATATTCAACT	600
TCAGCATCGA AGGGCAAGAT TGGGGTGTCT GGCCTAAGGA GACCCTCTAC AGCCCCAGCT	660
CGTGAAACAC CAAACGCCGG TGACATCAA ACCAGCGGAA AGCCCATCTT TTCTGCAGAA	720
GGCTATGCCA TTCAAGAAAC CAAATTCCTA AAAATCGAGG AATTGGACTT GGACTTCCAT	780
AACGAACCCA CGTTGAAGTT CCCCAAACCG GGTGGGTTG AGTGCCATCC GCAGAAATTA	840
CTGGTGAACG TCGTCCAATG CCTTGCCTCA AGTTTGCTCT CTCTGCAGAC TATCAACAGC	900
GAACGTGTAG CAAACGGTCT CCCACCTTAC AAGGTAATAT GCATGGGTAT AGCAAACATG	960
AGAGAAACCA CAATTCTGTG GTCCCGCCGC ACAGGAAAAC CAATTGTAA CTACGGTATT	1020
GTTTGAACG ACACCAGAAC GATCAAAATC GTTAGAGACA AATGGCAAAA CACTAGCGTC	1080
GATAGGCAAC TGCAGCTTAG ACAGAAGACT GGATTGCCAT TGCTCTCCAC GTATTTCTCC	1140
TGTTCCAAGC TGCCTGGTT CCTCGACAAT GAGCCTCTGT GTACCAAGGC GTATGAGGAG	1200
AACGACCTGA GTTTCGGCAC TGTGGACACA TGGCTGATTT ACCAATTAAC TAAACAAAAG	1260
GCGTTCGTTT CTGACGTAAC CAACGCTTCC AGAACTGGAT TTATGAACCT CTCCACTTTA	1320
AAGTACGACA ACGAGTTGCT GGAATTTTGG GGTATTGACA AGAACCTGAT TCACATGCCC	1380
GAAATTGTGT CCTCATCTCA ATACTACGGT GACTTTGGCA TTCCTGATTG GATAATGGAA	1440
AAGTACACG ATTCGCCAAA AACAGTACTG CGAGATCTAG TCAAGAGAAA CCTGCCCATA	1500
CAGGGCTGTC TGGGCGACCA AAGCGCATCC ATGGTGGGGC AACTCGCTTA CAAACCCGGT	1560
GCTGCAAAAT GTACTTATGG TACCGGTTGC TTTTACTGT ACAATACGGG GACCAAAAAA	1620
TTGATCTCCC AACATGGCGC ACTGACGACT CTAGCATTTT GGTTCACACA TTTGCAAGAG	1680
TACGGTGGCC AAAAACCAGA ATTGAGCAAG CCACATTTTG CATTAGAGGG TTCCGTCGCT	1740
GTGGCTGGTG CTGTGGTCCA ATGGCTACGT GATAATTTAC GATTGATCGA TAAATCAGAG	1800
GATGTCGGAC CGATTGCATC TACGGTTCCT GATTCTGGTG GCGTAGTTT CGTCCCCGCA	1860
TTTAGTGGCC TATTCGCTCC CTATTGGGAC CCAGATGCCA GAGCCACCAT AATGGGGATG	1920

TCTCAATTCA CTA CTACTGCCTC CCACATCGCC AGAGCTGCCG TGGAAGGTGT TTGCTTTCAA 1980
 GCCAGGGCTA TCTTGAAGGC AATGAGTTCT GACGCGTTTG GTGAAGGTTC CAAAGACAGG 2040
 GACTTTTTAG AGGAAATTC CGACGTCACA TATGAAAAGT CGCCCCTGTC GGTTCCTGGCA 2100
 GTGGATGGCG GGATGTCGAG GTCTAATGAA GTCATGCAAA TTCAAGCCGA TATCCTAGGT 2160
 CCCTGTGTCA AAGTCAGAAG GTCTCCGACA GCGGAATGTA CCGCATTGGG GGCAGCCATT 2220
 GCAGCCAATA TGGCTTTCAA GGATGTGAAC GAGCGCCCAT TATGGAAGGA CCTACACGAT 2280
 GTTAAGAAAT GGGTCTTTTA CAATGGAATG GAGAAAAACG AACAAATATC ACCAGAGGCT 2340
 CATCCAAACC TTAAGATATT CAGAAGTGAA TCCGACGATG CTGAAAGGAG AAAGCATTGG 2400
 AAGTATTGGG AAGTTGCCGT GGAAAGATCC AAAGGTTGGC TGAAGGACAT AGAAGGTGAA 2460
 CACGAACAGG TTCTAGAAAA CTTCCAATAA CAACATAAAT AATTTCTATT AACAAATGTAA 2520

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 391 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: unknown
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: GPD1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Met Ser Ala Ala Ala Asp Arg Leu Asn Leu Thr Ser Gly His Leu Asn
 1 5 10 15
 Ala Gly Arg Lys Arg Ser Ser Ser Val Ser Leu Lys Ala Ala Glu
 20 25 30
 Lys Pro Phe Lys Val Thr Val Ile Gly Ser Gly Asn Trp Gly Thr Thr
 35 40 45
 Ile Ala Lys Val Val Ala Glu Asn Cys Lys Gly Tyr Pro Glu Val Phe
 50 55 60
 Ala Pro Ile Val Gln Met Trp Val Phe Glu Glu Glu Ile Asn Gly Glu
 65 70 75 80
 Lys Leu Thr Glu Ile Ile Asn Thr Arg His Gln Asn Val Lys Tyr Leu
 85 90 95
 Pro Gly Ile Thr Leu Pro Asp Asn Leu Val Ala Asn Pro Asp Leu Ile
 100 105 110
 Asp Ser Val Lys Asp Val Asp Ile Ile Val Phe Asn Ile Pro His Gln
 115 120 125
 Phe Leu Pro Arg Ile Cys Ser Gln Leu Lys Gly His Val Asp Ser His
 130 135 140
 Val Arg Ala Ile Ser Cys Leu Lys Gly Phe Glu Val Gly Ala Lys Gly
 145 150 155 160

Val Gln Leu Leu Ser Ser Tyr Ile Thr Glu Glu Leu Gly Ile Gln Cys
 165 170 175
 Gly Ala Leu Ser Gly Ala Asn Ile Ala Thr Glu Val Ala Gln Glu His
 180 185 190
 Trp Ser Glu Thr Thr Val Ala Tyr His Ile Pro Lys Asp Phe Arg Gly
 195 200 205
 Glu Gly Lys Asp Val Asp His Lys Val Leu Lys Ala Leu Phe His Arg
 210 215 220
 Pro Tyr Phe His Val Ser Val Ile Glu Asp Val Ala Gly Ile Ser Ile
 225 230 235 240
 Cys Gly Ala Leu Lys Asn Val Val Ala Leu Gly Cys Gly Phe Val Glu
 245 250 255
 Gly Leu Gly Trp Gly Asn Asn Ala Ser Ala Ala Ile Gln Arg Val Gly
 260 265 270
 Leu Gly Glu Ile Ile Arg Phe Gly Gln Met Phe Phe Pro Glu Ser Arg
 275 280 285
 Glu Glu Thr Tyr Tyr Gln Glu Ser Ala Gly Val Ala Asp Leu Ile Thr
 290 295 300
 Thr Cys Ala Gly Gly Arg Asn Val Lys Val Ala Arg Leu Met Ala Thr
 305 310 315 320
 Ser Gly Lys Asp Ala Trp Glu Cys Glu Lys Glu Leu Leu Asn Gly Gln
 325 330 335
 Ser Ala Gln Gly Leu Ile Thr Cys Lys Glu Val His Glu Trp Leu Glu
 340 345 350
 Thr Cys Gly Ser Val Glu Asp Phe Pro Leu Phe Glu Ala Val Tyr Gln
 355 360 365
 Ile Val Tyr Asn Asn Tyr Pro Met Lys Asn Leu Pro Asp Met Ile Glu
 370 375 380
 Glu Leu Asp Leu His Glu Asp
 385 390

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 384 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: unknown
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: GPD2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met Thr Ala His Thr Asn Ile Lys Gln His Lys His Cys His Glu Asp
 1 5 10 15
 His Pro Ile Arg Arg Ser Asp Ser Ala Val Ser Ile Val His Leu Lys
 20 25 30

Arg Ala Pro Phe Lys Val Thr Val Ile Gly Ser Gly Asn Trp Gly Thr
 35 40 45
 Thr Ile Ala Lys Val Ile Ala Glu Asn Thr Glu Leu His Ser His Ile
 50 55 60
 Phe Glu Pro Glu Val Arg Met Trp Val Phe Asp Glu Lys Ile Gly Asp
 65 70 75 80
 Glu Asn Leu Thr Asp Ile Ile Asn Thr Arg His Gln Asn Val Lys Tyr
 85 90 95
 Leu Pro Asn Ile Asp Leu Pro His Asn Leu Val Ala Asp Pro Asp Leu
 100 105 110
 Leu His Ser Ile Lys Gly Ala Asp Ile Leu Val Phe Asn Ile Pro His
 115 120 125
 Gln Phe Leu Pro Asn Ile Val Lys Gln Leu Gln Gly His Val Ala Pro
 130 135 140
 His Val Arg Ala Ile Ser Cys Leu Lys Gly Phe Glu Leu Gly Ser Lys
 145 150 155 160
 Gly Val Gln Leu Leu Ser Ser Tyr Val Thr Asp Glu Leu Gly Ile Gln
 165 170 175
 Cys Gly Ala Leu Ser Gly Ala Asn Leu Ala Pro Glu Val Ala Lys Glu
 180 185 190
 His Trp Ser Glu Thr Thr Val Ala Tyr Gln Leu Pro Lys Asp Tyr Gln
 195 200 205
 Gly Asp Gly Lys Asp Val Asp His Lys Ile Leu Lys Leu Leu Phe His
 210 215 220
 Arg Pro Tyr Phe His Val Asn Val Ile Asp Asp Val Ala Gly Ile Ser
 225 230 235 240
 Ile Ala Gly Ala Leu Lys Asn Val Val Ala Leu Ala Cys Gly Phe Val
 245 250 255
 Glu Gly Met Gly Trp Gly Asn Asn Ala Ser Ala Ala Ile Gln Arg Leu
 260 265 270
 Gly Leu Gly Glu Ile Ile Lys Phe Gly Arg Met Phe Phe Pro Glu Ser
 275 280 285
 Lys Val Glu Thr Tyr Tyr Gln Glu Ser Ala Gly Val Ala Asp Leu Ile
 290 295 300
 Thr Thr Cys Ser Gly Gly Arg Asn Val Lys Val Ala Thr Tyr Met Ala
 305 310 315 320
 Lys Thr Gly Lys Ser Ala Leu Glu Ala Glu Lys Glu Leu Leu Asn Gly
 325 330 335
 Gln Ser Ala Gln Gly Ile Ile Thr Cys Arg Glu Val His Glu Trp Leu
 340 345 350
 Gln Thr Cys Glu Leu Thr Gln Glu Phe Pro Ile Ile Arg Gly Ser Leu
 355 360 365
 Pro Asp Ser Leu Gln Gln Arg Pro His Gly Arg Pro Thr Gly Asp Asp
 370 375 380

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 614 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: unknown
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: GUT2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```

Met Thr Arg Ala Thr Trp Cys Asn Ser Pro Pro Pro Leu His Arg Gln
1          5          10          15
Val Ser Arg Arg Asp Leu Leu Asp Arg Leu Asp Lys Thr His Gln Phe
20          25          30
Asp Val Leu Ile Ile Gly Gly Gly Ala Thr Gly Thr Gly Cys Ala Leu
35          40          45
Asp Ala Ala Thr Arg Gly Leu Asn Val Ala Leu Val Glu Lys Gly Asp
50          55          60
Phe Ala Ser Gly Thr Ser Ser Lys Ser Thr Lys Met Ile His Gly Gly
65          70          75          80
Val Arg Tyr Leu Glu Lys Ala Phe Trp Glu Phe Ser Lys Ala Gln Leu
85          90          95
Asp Leu Val Ile Glu Ala Leu Asn Glu Arg Lys His Leu Ile Asn Thr
100         105         110
Ala Pro His Leu Cys Thr Val Leu Pro Ile Leu Ile Pro Ile Tyr Ser
115         120         125
Thr Trp Gln Val Pro Tyr Ile Tyr Met Gly Cys Lys Phe Tyr Asp Phe
130         135         140
Phe Gly Gly Ser Gln Asn Leu Lys Lys Ser Tyr Leu Leu Ser Lys Ser
145         150         155         160
Ala Thr Val Glu Lys Ala Pro Met Leu Thr Thr Asp Asn Leu Lys Ala
165         170         175
Ser Leu Val Tyr His Asp Gly Ser Phe Asn Asp Ser Arg Leu Asn Ala
180         185         190
Thr Leu Ala Ile Thr Gly Val Glu Asn Gly Ala Thr Val Leu Ile Tyr
195         200         205
Val Glu Val Gln Lys Leu Ile Lys Asp Pro Thr Ser Gly Lys Val Ile
210         215         220
Gly Ala Glu Ala Arg Asp Val Glu Thr Asn Glu Leu Val Arg Ile Asn
225         230         235         240
Ala Lys Cys Val Val Asn Ala Thr Gly Pro Tyr Ser Asp Ala Ile Leu
245         250         255
Gln Met Asp Arg Asn Pro Ser Gly Leu Pro Asp Ser Pro Leu Asn Asp
260         265         270

```

Asn Ser Lys Ile Lys Ser Thr Phe Asn Gln Ile Ser Val Met Asp Pro
 275 280 285
 Lys Met Val Ile Pro Ser Ile Gly Val His Ile Val Leu Pro Ser Phe
 290 295 300
 Tyr Ser Pro Lys Asp Met Gly Leu Leu Asp Val Arg Thr Ser Asp Gly
 305 310 315 320
 Arg Val Met Phe Phe Leu Pro Trp Gln Gly Lys Val Leu Ala Gly Thr
 325 330 335
 Thr Asp Ile Pro Leu Lys Gln Val Pro Glu Asn Pro Met Pro Thr Glu
 340 345 350
 Ala Asp Ile Gln Asp Ile Leu Lys Glu Leu Gln His Tyr Ile Glu Phe
 355 360 365
 Pro Val Lys Arg Glu Asp Val Leu Ser Ala Trp Ala Gly Val Arg Pro
 370 375 380
 Leu Val Arg Asp Pro Arg Thr Ile Pro Ala Asp Gly Lys Lys Gly Ser
 385 390 395 400
 Ala Thr Gln Gly Val Val Arg Ser His Phe Leu Phe Thr Ser Asp Asn
 405 410 415
 Gly Leu Ile Thr Ile Ala Gly Gly Lys Trp Thr Thr Tyr Arg Gln Met
 420 425 430
 Ala Glu Glu Thr Val Asp Lys Val Val Glu Val Gly Gly Phe His Asn
 435 440 445
 Leu Lys Pro Cys His Thr Arg Asp Ile Lys Leu Ala Gly Ala Glu Glu
 450 455 460
 Trp Thr Gln Asn Tyr Val Ala Leu Leu Ala Gln Asn Tyr His Leu Ser
 465 470 475 480
 Ser Lys Met Ser Asn Tyr Leu Val Gln Asn Tyr Gly Thr Arg Ser Ser
 485 490 495
 Ile Ile Cys Glu Phe Phe Lys Glu Ser Met Glu Asn Lys Leu Pro Leu
 500 505 510
 Ser Leu Ala Asp Lys Glu Asn Asn Val Ile Tyr Ser Ser Glu Glu Asn
 515 520 525
 Asn Leu Val Asn Phe Asp Thr Phe Arg Tyr Pro Phe Thr Ile Gly Glu
 530 535 540
 Leu Lys Tyr Ser Met Gln Tyr Glu Tyr Cys Arg Thr Pro Leu Asp Phe
 545 550 555 560
 Leu Leu Arg Arg Thr Arg Phe Ala Phe Leu Asp Ala Lys Glu Ala Leu
 565 570 575
 Asn Ala Val His Ala Thr Val Lys Val Met Gly Asp Glu Phe Asn Trp
 580 585 590
 Ser Glu Lys Lys Arg Gln Trp Glu Leu Glu Lys Thr Val Asn Phe Ile
 595 600 605
 Gln Gly Arg Phe Gly Val
 610

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 339 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: GPSA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```

Met Asn Gln Arg Asn Ala Ser Met Thr Val Ile Gly Ala Gly Ser Tyr
1           5           10           15
Gly Thr Ala Leu Ala Ile Thr Leu Ala Arg Asn Gly His Glu Val Val
20           25           30
Leu Trp Gly His Asp Pro Glu His Ile Ala Thr Leu Glu Arg Asp Arg
35           40           45
Cys Asn Ala Ala Phe Leu Pro Asp Val Pro Phe Pro Asp Thr Leu His
50           55           60
Leu Glu Ser Asp Leu Ala Thr Ala Leu Ala Ala Ser Arg Asn Ile Leu
65           70           75           80
Val Val Val Pro Ser His Val Phe Gly Glu Val Leu Arg Gln Ile Lys
85           90           95
Pro Leu Met Arg Pro Asp Ala Arg Leu Val Trp Ala Thr Lys Gly Leu
100          105          110
Glu Ala Glu Thr Gly Arg Leu Leu Gln Asp Val Ala Arg Glu Ala Leu
115          120          125
Gly Asp Gln Ile Pro Leu Ala Val Ile Ser Gly Pro Thr Phe Ala Lys
130          135          140
Glu Leu Ala Ala Gly Leu Pro Thr Ala Ile Ser Leu Ala Ser Thr Asp
145          150          155          160
Gln Thr Phe Ala Asp Asp Leu Gln Gln Leu Leu His Cys Gly Lys Ser
165          170          175
Phe Arg Val Tyr Ser Asn Pro Asp Phe Ile Gly Val Gln Leu Gly Gly
180          185          190
Ala Val Lys Asn Val Ile Ala Ile Gly Ala Gly Met Ser Asp Gly Ile
195          200          205
Gly Phe Gly Ala Asn Ala Arg Thr Ala Leu Ile Thr Arg Gly Leu Ala
210          215          220
Glu Met Ser Arg Leu Gly Ala Ala Leu Gly Ala Asp Pro Ala Thr Phe
225          230          235          240
Met Gly Met Ala Gly Leu Gly Asp Leu Val Leu Thr Cys Thr Asp Asn
245          250          255
Gln Ser Arg Asn Arg Arg Phe Gly Met Met Leu Gly Gln Gly Met Asp
260          265          270

```

Val Gln Ser Ala Gln Glu Lys Ile Gly Gln Val Val Glu Gly Tyr Arg
 275 280 285

Asn Thr Lys Glu Val Arg Glu Leu Ala His Arg Phe Gly Val Glu Met
 290 295 300

Pro Ile Thr Glu Glu Ile Tyr Gln Val Leu Tyr Cys Gly Lys Asn Ala
 305 310 315 320

Arg Glu Ala Ala Leu Thr Leu Leu Gly Arg Ala Arg Lys Asp Glu Arg
 325 330 335

Ser Ser His

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 501 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: unknown
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: GLPD

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Met Glu Thr Lys Asp Leu Ile Val Ile Gly Gly Gly Ile Asn Gly Ala
 1 5 10 15

Gly Ile Ala Ala Asp Ala Ala Gly Arg Gly Leu Ser Val Leu Met Leu
 20 25 30

Glu Ala Gln Asp Leu Ala Cys Ala Thr Ser Ser Ala Ser Ser Lys Leu
 35 40 45

Ile His Gly Gly Leu Arg Tyr Leu Glu His Tyr Glu Phe Arg Leu Val
 50 55 60

Ser Glu Ala Leu Ala Glu Arg Glu Val Leu Leu Lys Met Ala Pro His
 65 70 75 80

Ile Ala Phe Pro Met Arg Phe Arg Leu Pro His Arg Pro His Leu Arg
 85 90 95

Pro Ala Trp Met Ile Arg Ile Gly Leu Phe Met Tyr Asp His Leu Gly
 100 105 110

Lys Arg Thr Ser Leu Pro Gly Ser Thr Gly Leu Arg Phe Gly Ala Asn
 115 120 125

Ser Val Leu Lys Pro Glu Ile Lys Arg Gly Phe Glu Tyr Ser Asp Cys
 130 135 140

Trp Val Asp Asp Ala Arg Leu Val Leu Ala Asn Ala Gln Met Val Val
 145 150 155 160

Arg Lys Gly Gly Glu Val Leu Thr Arg Thr Arg Ala Thr Ser Ala Arg
 165 170 175

Arg Glu Asn Gly Leu Trp Ile Val Glu Ala Glu Asp Ile Asp Thr Gly
 180 185 190

Lys Lys Tyr Ser Trp Gln Ala Arg Gly Leu Val Asn Ala Thr Gly Pro
 195 200 205
 Trp Val Lys Gln Phe Phe Asp Asp Gly Met His Leu Pro Ser Pro Tyr
 210 215 220
 Gly Ile Arg Leu Ile Lys Gly Ser His Ile Val Val Pro Arg Val His
 225 230 235 240
 Thr Gln Lys Gln Ala Tyr Ile Leu Gln Asn Glu Asp Lys Arg Ile Val
 245 250 255
 Phe Val Ile Pro Trp Met Asp Glu Phe Ser Ile Ile Gly Thr Thr Asp
 260 265 270
 Val Glu Tyr Lys Gly Asp Pro Lys Ala Val Lys Ile Glu Glu Ser Glu
 275 280 285
 Ile Asn Tyr Leu Leu Asn Val Tyr Asn Thr His Phe Lys Lys Gln Leu
 290 295 300
 Ser Arg Asp Asp Ile Val Trp Thr Tyr Ser Gly Val Arg Pro Leu Cys
 305 310 315 320
 Asp Asp Glu Ser Asp Ser Pro Gln Ala Ile Thr Arg Asp Tyr Thr Leu
 325 330 335
 Asp Ile His Asp Glu Asn Gly Lys Ala Pro Leu Leu Ser Val Phe Gly
 340 345 350
 Gly Lys Leu Thr Thr Tyr Arg Lys Leu Ala Glu His Ala Leu Glu Lys
 355 360 365
 Leu Thr Pro Tyr Tyr Gln Gly Ile Gly Pro Ala Trp Thr Lys Glu Ser
 370 375 380
 Val Leu Pro Gly Gly Ala Ile Glu Gly Asp Arg Asp Asp Tyr Ala Ala
 385 390 395 400
 Arg Leu Arg Arg Arg Tyr Pro Phe Leu Thr Glu Ser Leu Ala Arg His
 405 410 415
 Tyr Ala Arg Thr Tyr Gly Ser Asn Ser Glu Leu Leu Leu Gly Asn Ala
 420 425 430
 Gly Thr Val Ser Asp Leu Gly Glu Asp Phe Gly His Glu Phe Tyr Glu
 435 440 445
 Ala Glu Leu Lys Tyr Leu Val Asp His Glu Trp Val Arg Arg Ala Asp
 450 455 460
 Asp Ala Leu Trp Arg Arg Thr Lys Gln Gly Met Trp Leu Asn Ala Asp
 465 470 475 480
 Gln Gln Ser Arg Val Ser Gln Trp Leu Val Glu Tyr Thr Gln Gln Arg
 485 490 495
 Leu Ser Leu Ala Ser
 500

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 542 amino acids
 - (B) TYPE: amino acid

(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: GLPABC

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```

Met Lys Thr Arg Asp Ser Gln Ser Ser Asp Val Ile Ile Ile Gly Gly
1      5      10      15
Gly Ala Thr Gly Ala Gly Ile Ala Arg Asp Cys Ala Leu Arg Gly Leu
20      25      30
Arg Val Ile Leu Val Glu Arg His Asp Ile Ala Thr Gly Ala Thr Gly
35      40      45
Arg Asn His Gly Leu Leu His Ser Gly Ala Arg Tyr Ala Val Thr Asp
50      55      60
Ala Glu Ser Ala Arg Glu Cys Ile Ser Glu Asn Gln Ile Leu Lys Arg
65      70      75      80
Ile Ala Arg His Cys Val Glu Pro Thr Asn Gly Leu Phe Ile Thr Leu
85      90      95
Pro Glu Asp Asp Leu Ser Phe Gln Ala Thr Phe Ile Arg Ala Cys Glu
100     105     110
Glu Ala Gly Ile Ser Ala Glu Ala Ile Asp Pro Gln Gln Ala Arg Ile
115     120     125
Ile Glu Pro Ala Val Asn Pro Ala Leu Ile Gly Ala Val Lys Val Pro
130     135     140
Asp Gly Thr Val Asp Pro Phe Arg Leu Thr Ala Ala Asn Met Leu Asp
145     150     155     160
Ala Lys Glu His Gly Ala Val Ile Leu Thr Ala His Glu Val Thr Gly
165     170     175
Leu Ile Arg Glu Gly Ala Thr Val Cys Gly Val Arg Val Arg Asn His
180     185     190
Leu Thr Gly Glu Thr Gln Ala Leu His Ala Pro Val Val Val Asn Ala
195     200     205
Ala Gly Ile Trp Gly Gln His Ile Ala Glu Tyr Ala Asp Leu Arg Ile
210     215     220
Arg Met Phe Pro Ala Lys Gly Ser Leu Leu Ile Met Asp His Arg Ile
225     230     235     240
Asn Gln His Val Ile Asn Arg Cys Arg Lys Pro Ser Asp Ala Asp Ile
245     250     255
Leu Val Pro Gly Asp Thr Ile Ser Leu Ile Gly Thr Thr Ser Leu Arg
260     265     270
Ile Asp Tyr Asn Glu Ile Asp Asp Asn Arg Val Thr Ala Glu Glu Val
275     280     285
Asp Ile Leu Leu Arg Glu Gly Glu Lys Leu Ala Pro Val Met Ala Lys
290     295     300

```


Thr Arg Ile Leu Arg Ala Tyr Ser Gly Val Arg Pro Leu Val Ala Ser
 305 310 315 320
 Asp Asp Asp Pro Ser Gly Arg Asn Leu Ser Arg Gly Ile Val Leu Leu
 325 330 335
 Asp His Ala Glu Arg Asp Gly Leu Asp Gly Phe Ile Thr Ile Thr Gly
 340 345 350
 Gly Lys Leu Met Thr Tyr Arg Leu Met Ala Glu Trp Ala Thr Asp Ala
 355 360 365
 Val Cys Arg Lys Leu Gly Asn Thr Arg Pro Cys Thr Thr Ala Asp Leu
 370 375 380
 Ala Leu Pro Gly Ser Gln Glu Pro Ala Glu Val Thr Leu Arg Lys Val
 385 390 395 400
 Ile Ser Leu Pro Ala Pro Leu Arg Gly Ser Ala Val Tyr Arg His Gly
 405 410 415
 Asp Arg Thr Pro Ala Trp Leu Ser Glu Gly Arg Leu His Arg Ser Leu
 420 425 430
 Val Cys Glu Cys Glu Ala Val Thr Ala Gly Glu Val Gln Tyr Ala Val
 435 440 445
 Glu Asn Leu Asn Val Asn Ser Leu Leu Asp Leu Arg Arg Arg Thr Arg
 450 455 460
 Val Gly Met Gly Thr Cys Gln Gly Glu Leu Cys Ala Cys Arg Ala Ala
 465 470 475 480
 Gly Leu Leu Gln Arg Phe Asn Val Thr Thr Ser Ala Gln Ser Ile Glu
 485 490 495
 Gln Leu Ser Thr Phe Leu Asn Glu Arg Trp Lys Gly Val Gln Pro Ile
 500 505 510
 Ala Trp Gly Asp Ala Leu Arg Glu Ser Glu Phe Thr Arg Trp Val Tyr
 515 520 525
 Gln Gly Leu Cys Gly Leu Glu Lys Glu Gln Lys Asp Ala Leu
 530 535 540

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 250 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: unknown
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: GPP2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Met Gly Leu Thr Thr Lys Pro Leu Ser Leu Lys Val Asn Ala Ala Leu
 1 5 10 15
 Phe Asp Val Asp Gly Thr Ile Ile Ile Ser Gln Pro Ala Ile Ala Ala
 20 25 30

Phe Trp Arg Asp Phe Gly Lys Asp Lys Pro Tyr Phe Asp Ala Glu His
 35 40 45
 Val Ile Gln Val Ser His Gly Trp Arg Thr Phe Asp Ala Ile Ala Lys
 50 55 60
 Phe Ala Pro Asp Phe Ala Asn Glu Glu Tyr Val Asn Lys Leu Glu Ala
 65 70 75 80
 Glu Ile Pro Val Lys Tyr Gly Glu Lys Ser Ile Glu Val Pro Gly Ala
 85 90 95
 Val Lys Leu Cys Asn Ala Leu Asn Ala Leu Pro Lys Glu Lys Trp Ala
 100 105 110
 Val Ala Thr Ser Gly Thr Arg Asp Met Ala Gln Lys Trp Phe Glu His
 115 120 125
 Leu Gly Ile Arg Arg Pro Lys Tyr Phe Ile Thr Ala Asn Asp Val Lys
 130 135 140
 Gln Gly Lys Pro His Pro Glu Pro Tyr Leu Lys Gly Arg Asn Gly Leu
 145 150 155 160
 Gly Tyr Pro Ile Asn Glu Gln Asp Pro Ser Lys Ser Lys Val Val Val
 165 170 175
 Phe Glu Asp Ala Pro Ala Gly Ile Ala Ala Gly Lys Ala Ala Gly Cys
 180 185 190
 Lys Ile Ile Gly Ile Ala Thr Thr Phe Asp Leu Asp Phe Leu Lys Glu
 195 200 205
 Lys Gly Cys Asp Ile Ile Val Lys Asn His Glu Ser Ile Arg Val Gly
 210 215 220
 Gly Tyr Asn Ala Glu Thr Asp Glu Val Glu Phe Ile Phe Asp Asp Tyr
 225 230 235 240
 Leu Tyr Ala Lys Asp Asp Leu Leu Lys Trp
 245 250

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 709 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: GUT1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Met Phe Pro Ser Leu Phe Arg Leu Val Val Phe Ser Lys Arg Tyr Ile
 1 5 10 15
 Phe Arg Ser Ser Gln Arg Leu Tyr Thr Ser Leu Lys Gln Glu Gln Ser
 20 25 30
 Arg Met Ser Lys Ile Met Glu Asp Leu Arg Ser Asp Tyr Val Pro Leu
 35 40 45

Ile Ala Ser Ile Asp Val Gly Thr Thr Ser Ser Arg Cys Ile Leu Phe
 50 55 60
 Asn Arg Trp Gly Gln Asp Val Ser Lys His Gln Ile Glu Tyr Ser Thr
 65 70 75 80
 Ser Ala Ser Lys Gly Lys Ile Gly Val Ser Gly Leu Arg Arg Pro Ser
 85 90 95
 Thr Ala Pro Ala Arg Glu Thr Pro Asn Ala Gly Asp Ile Lys Thr Ser
 100 105 110
 Gly Lys Pro Ile Phe Ser Ala Glu Gly Tyr Ala Ile Gln Glu Thr Lys
 115 120 125
 Phe Leu Lys Ile Glu Glu Leu Asp Leu Asp Phe His Asn Glu Pro Thr
 130 135 140
 Leu Lys Phe Pro Lys Pro Gly Trp Val Glu Cys His Pro Gln Lys Leu
 145 150 155 160
 Leu Val Asn Val Val Gln Cys Leu Ala Ser Ser Leu Leu Ser Leu Gln
 165 170 175
 Thr Ile Asn Ser Glu Arg Val Ala Asn Gly Leu Pro Pro Tyr Lys Val
 180 185 190
 Ile Cys Met Gly Ile Ala Asn Met Arg Glu Thr Thr Ile Leu Trp Ser
 195 200 205
 Arg Arg Thr Gly Lys Pro Ile Val Asn Tyr Gly Ile Val Trp Asn Asp
 210 215 220
 Thr Arg Thr Ile Lys Ile Val Arg Asp Lys Trp Gln Asn Thr Ser Val
 225 230 235 240
 Asp Arg Gln Leu Gln Leu Arg Gln Lys Thr Gly Leu Pro Leu Leu Ser
 245 250 255
 Thr Tyr Phe Ser Cys Ser Lys Leu Arg Trp Phe Leu Asp Asn Glu Pro
 260 265 270
 Leu Cys Thr Lys Ala Tyr Glu Glu Asn Asp Leu Met Phe Gly Thr Val
 275 280 285
 Asp Thr Trp Leu Ile Tyr Gln Leu Thr Lys Gln Lys Ala Phe Val Ser
 290 295 300
 Asp Val Thr Asn Ala Ser Arg Thr Gly Phe Met Asn Leu Ser Thr Leu
 305 310 315 320
 Lys Tyr Asp Asn Glu Leu Leu Glu Phe Trp Gly Ile Asp Lys Asn Leu
 325 330 335
 Ile His Met Pro Glu Ile Val Ser Ser Ser Gln Tyr Tyr Gly Asp Phe
 340 345 350
 Gly Ile Pro Asp Trp Ile Met Glu Lys Leu His Asp Ser Pro Lys Thr
 355 360 365
 Val Leu Arg Asp Leu Val Lys Arg Asn Leu Pro Ile Gln Gly Cys Leu
 370 375 380
 Gly Asp Gln Ser Ala Ser Met Val Gly Gln Leu Ala Tyr Lys Pro Gly
 385 390 395 400

Ala Ala Lys Cys Thr Tyr Gly Thr Gly Cys Phe Leu Leu Tyr Asn Thr
 405 410 415
 Gly Thr Lys Lys Leu Ile Ser Gln His Gly Ala Leu Thr Thr Leu Ala
 420 425 430
 Phe Trp Phe Pro His Leu Gln Glu Tyr Gly Gly Gln Lys Pro Glu Leu
 435 440 445
 Ser Lys Pro His Phe Ala Leu Glu Gly Ser Val Ala Val Ala Gly Ala
 450 455 460
 Val Val Gln Trp Leu Arg Asp Asn Leu Arg Leu Ile Asp Lys Ser Glu
 465 470 475 480
 Asp Val Gly Pro Ile Ala Ser Thr Val Pro Asp Ser Gly Gly Val Val
 485 490 495
 Phe Val Pro Ala Phe Ser Gly Leu Phe Ala Pro Tyr Trp Asp Pro Asp
 500 505 510
 Ala Arg Ala Thr Ile Met Gly Met Ser Gln Phe Thr Thr Ala Ser His
 515 520 525
 Ile Ala Arg Ala Ala Val Glu Gly Val Cys Phe Gln Ala Arg Ala Ile
 530 535 540
 Leu Lys Ala Met Ser Ser Asp Ala Phe Gly Glu Gly Ser Lys Asp Arg
 545 550 555 560
 Asp Phe Leu Glu Glu Ile Ser Asp Val Thr Tyr Glu Lys Ser Pro Leu
 565 570 575
 Ser Val Leu Ala Val Asp Gly Gly Met Ser Arg Ser Asn Glu Val Met
 580 585 590
 Gln Ile Gln Ala Asp Ile Leu Gly Pro Cys Val Lys Val Arg Arg Ser
 595 600 605
 Pro Thr Ala Glu Cys Thr Ala Leu Gly Ala Ala Ile Ala Ala Asn Met
 610 615 620
 Ala Phe Lys Asp Val Asn Glu Arg Pro Leu Trp Lys Asp Leu His Asp
 625 630 635 640
 Val Lys Lys Trp Val Phe Tyr Asn Gly Met Glu Lys Asn Glu Gln Ile
 645 650 655
 Ser Pro Glu Ala His Pro Asn Leu Lys Ile Phe Arg Ser Glu Ser Asp
 660 665 670
 Asp Ala Glu Arg Arg Lys His Trp Lys Tyr Trp Glu Val Ala Val Glu
 675 680 685
 Arg Ser Lys Gly Trp Leu Lys Asp Ile Glu Gly Glu His Glu Gln Val
 690 695 700
 Leu Glu Asn Phe Gln
 705

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 12145 base pairs
 - (B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:
(A) ORGANISM: PHK28-26

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

```

GTCGACCACC ACGGTGGTGA CTTTAATGCC GCTCTCATGC AGCAGCTCGG TGGCGGTCTC      60
AAAATTTCAGG ATGTCGCCGG TATAGTTTTT GATAATCAGC AAGACGCCTT CGCCGCCGTC      120
AATTTGCATC GCGCATTCAA ACATTTTGTC CGGCGTCGGC GAGGTGAATA TTTCCCCCGG      180
ACAGGCGCCG GAGAGCATGC CCTGGCCGAT ATAGCCGCAG TGCATCGGTT CATGTCCGCT      240
GCCGCCGCCG GAGAGCAGGG CCACCTTGCC AGCCACCGGC GCGTCGGTGC GGGTCACATA      300
CAGCGGGTCC TGATGCAGGG TCAGCTGCGG ATGGGCTTTA GCCAGCCCCT GTAATTGTTC      360
ATTCAGTACA TCTTCAACAC GGTTAATCAG CTTTTTCATT ATTCAGTGCT CCGTTGGAGA      420
AGGTTTCGATG CCGCCTCTCT GCTGGCGGAG GCGGTCATCG CGTAGGGGTA TCGTCTGACG      480
GTGGAGCGTG CCTGGCGATA TGATGATTCT GGCTGAGCGG ACGAAAAAAA GAATGCCCCG      540
ACGATCGGGT TTCATTACGA AACATTGCTT CCTGATTTTG TTTCTTTATG GAACGTTTTT      600
GCTGAGGATA TGGTGAAAAT GCGAGCTGGC GCGCTTTTTT TCTTCTGCCA TAAGCGGCGG      660
TCAGGATAGC CGGCGAAGCG GGTGGGAAAA AATTTTTTGC TGATTTTCTG CCGACTGCGG      720
GAGAAAAGGC GGTCAAACAC GGAGGATTGT AAGGGCATTG TCGCGCAAAG GAGCGGATCG      780
GGATCGCAAT CCTGACAGAG ACTAGGGTTT TTTGTTCCAA TATGGAACGT AAAAAATTAA      840
CCTGTGTTTC ATATCAGAAC AAAAAGGCGA AAGATTTTTT TGTTCCCTGC CGGCCCTACA      900
GTGATCGCAC TGCTCCGTA CGCTCCGTTC AGGCCGCGCT TCACTGGCCG GCGCGGATAA      960
CGCCAGGGCT CATCATGTCT ACATGCGCAC TTATTTGAGG GTGAAAGGAA TGCTAAAAGT     1020
TATTCAATCT CCAGCCAAAT ATCTTCAGGG TCCTGATGCT GCTGTTCTGT TCGGTCAATA     1080
TGCCAAAAAC CTGGCGGAGA GCTTCTTCGT CATCGCTGAC GATTTCGTAA TGAAGCTGGC     1140
GGGAGAGAAA GTGGTGAATG GCCTGCAGAG CCACGATATT CGCTGCCATG CGGAACGGTT     1200
TAACGGCGAA TGCAGCCATG CGGAAATCAA CCGTCTGATG GCGATTTTGC AAAAACAGGG     1260
CTGCCGCGGC GTGGTCGGGA TCGGCGGTGG TAAACCCCTC GATACCGCGA AGGCGATCGG     1320
TTACTACCAG AAGCTGCCGG TGGTGGTGAT CCCGACCATC GCCTCGACCG ATGCGCCAAC     1380
CAGCGCGCTG TCGGTGATCT ACACCGAAGC GGGCGAGTTT GAAGAGTATC TGATCTATCC     1440
GAAAAACCCG GATATGGTGG TGATGGACAC GGCGATTATC GCCAAAGCGC CGGTACGCCT     1500
GCTGGTCTCC GGCATGGGCG ATGCGCTCTC CACCTGGTTC GAGGCCAAAG CTTGCTACGA     1560
TGCGCGCGCC ACCAGCATGG CCGGAGGACA GTCCACCAG GCGGCGCTGA GCCTCGCCCG     1620
CCTGTGCTAT GATACGCTGC TGGCGGAGGG CGAAAAGGCC CGTCTGGCGG CGCAGGCCCG     1680

```

GGTAGTGACC GAAGCGCTGG AGCGCATCAT CGAGGCGAAC ACTTACCTCA GCGGCATTGG	1740
CTTTGAAAGC AGTGGCCTGG CCGCTGCCCA TGCAATCCAC AACGGTTTCA CCATTCTTGA	1800
AGAGTGCCAT CACCTGTATC ACGGTGAGAA AGTGGCCTTC GGTACCCTGG CGCAGCTGGT	1860
GCTGCAGAAC AGCCCAGATGG ACGAGATTGA AACGGTGCAG GGCTTCTGCC AGCGCGTCGG	1920
CCTGCCGGTG ACGCTCGCGC AGATGGGCGT CAAAGAGGGG ATCGACGAGA AAATCGCCGC	1980
GGTGGCGAAA GCTACCTGCG CGGAAGGGGA AACCATCCAT AATATGCCGT TTGCGGTGAC	2040
CCCGGAGAGC GTCCATGCCG CTATCCTCAC CGCCGATCTG TTAGGCCAGC AGTGGCTGGC	2100
GCGTTAATTC GCGGTGGCTA AACCGCTGGC CCAGGTCAGC GGTTTTTCTT TCTCCCCTCC	2160
GGCAGTCGCT GCCGGAGGGG TTCTCTATGG TACAACGCGG AAAAGGATAT GACTGTTTCA	2220
ACTCAGGATA CCGGGAAGGC GGTCTCTTCC GTCATTGCCC AGTCATGGCA CCGCTGCAGC	2280
AAGTTTATGC AGCGCGAAAC CTGGCAAACG CCGCACCAGG CCCAGGGCCT GACCTTCGAC	2340
TCCATCTGTC GCGGTAAGC CGCGCTGCTC ACCATCGGCC AGCGGCGCT GGAAGACGCC	2400
TGGGAGTTTA TGGACGGCCG CCCCTGCGCG CTGTTTATTC TTGATGAGTC CGCCTGCATC	2460
CTGAGCCGTT GCGGCGAGCC GCAAACCCTG GCCCAGCTGG CTGCCCTGGG ATTTGCGGAC	2520
GGCAGCTATT GTGCGGAGAG CATTATCGGC ACCTGCGCGC TGTCGCTGGC CGCGATGCAG	2580
GGCCAGCCGA TCAACACCGC CGGCGATCGG CATTTTAAGC AGGCGCTACA GCCATGGAGT	2640
TTTGTCTCGA CGCCGGTGTT TGATAACCAC GGGCGGCTGT TCGGCTCTAT CTCGCTTTGC	2700
TGTCTGGTCG AGCACCAGTC CAGCGCCGAC CTCTCCCTGA CGCTGGCCAT CGCCCGCGAG	2760
GTGGGTAACT CCCTGCTTAC CGACAGCCTG CTGGCGGAAT CCAACCGTCA CCTCAATCAG	2820
ATGTACGGCC TGCTGGAGAG CATGGACGAT GGGGTGATGG CGTGGAACGA ACAGGGCGTG	2880
CTGCAGTTTC TCAATGTTCA GCGGCGGAGA CTGCTGCATC TTGATGCTCA GGCCAGCCAG	2940
GGGAAAATA TCGCCGATCT GGTGACCCTC CCGGCGCTGC TCGCCCGCGC CATCAAACAC	3000
GCCCGCGGCC TGAATCACGT CGAAGTCACC TTTGAAAGTC AGCATCAGTT TGTGATGCG	3060
GTGATCACCT TAAAACCGAT TGTCGAGGCG CAAGGCAACA GTTTTATTCT GCTGCTGCAT	3120
CCGGTGGAGC AGATGCGGCA GCTGATGACC AGCCAGCTCG GTAAAGTCAG CCACACCTTT	3180
GAGCAGATGT CTGCCGACGA TCCGGAACC CGACGCTGA TCCACTTTGG CCGCCAGGCG	3240
GCGCGCGGCG GCTTCCCGGT GCTACTGTGC GGCGAAGAGG GGGTCGGGAA AGAGCTGCTG	3300
AGCCAGGCTA TTCACAATGA AAGCGAACGG GCGGGCGGCC CCTACATCTC CGTCAACTGC	3360
CAGCTATATG CCGACAGCGT GCTGGGCCAG GACTTTATGG GCAGCGCCCC TACCGACGAT	3420
GAAAATGGTC GCCTGAGCCG CCTTGAGCTG GCCAACGGCG GCACCCTGTT TCTGGAAAAG	3480
ATCGAGTATC TGGCGCCGGA GCTGCAGTCG GCTCTGCTGC AGGTGATTAA GCAGGGCGTG	3540
CTCACC CGCC TCGACGCCG GCGCCTGATC CCGGTGGATG TGAAGGTGAT TGCCACCACC	3600
ACCGTCGATC TGGCCAATCT GGTGGAACAG AACCGCTTTA GCCGCCAGCT GTACTATGCG	3660

CTGCACTCCT TTGAGATCGT CATCCCGCCG CTGCGCGCCC GACGCAACAG TATTCCGTCTG	3720
CTGGTGCCATA ACCGGTTGAA GAGCCTGGAG AAGCGTTTCT CTTCGCGACT GAAAGTGGAC	3780
GATGACGCGC TGGCACAGCT GGTGGCCTAC TCGTGGCCGG GGAATGATTT TGAGCTCAAC	3840
AGCGTCATTG AGAATATCGC CATCAGCAGC GACAACGGCC ACATTGCGCT GAGTAATCTG	3900
CCGGAATATC TCTTTTCCGA GCGGCCGGGC GGGGATAGCG CGTCATCGCT GCTGCCGGCC	3960
AGCCTGACTT TTAGCGCCAT CGAAAAGGAA GCTATTATTC ACGCCGCCCG GGTGACCAGC	4020
GGGCGGGTGC AGGAGATGTC GCAGCTGCTC AATATCGGCC GCACCACCCT GTGGCGCAAA	4080
ATGAAGCAGT ACGATATTGA CGCCAGCCAG TTCAAGCGCA AGCATCAGGC CTAGTCTCTT	4140
CGATTGCGCG CATGGAGAAC AGGGCATCCG ACAGGCGATT GCTGTAGCGT TTGAGCGCGT	4200
CGCGCAGCGG ATGCGCGCGG TCCATGGCCG TCAGCAGGCG TTCGAGCCGA CGGGACTGGG	4260
TGCGCGCCAC GTGCAGCTGG GCAGAGGCGA GATTCCCTCC CGGGATCACG AACTGTTTTA	4320
ACGGGCCGCT CTCGGCCATA TTGCGGTGCA TAAGCCGCTC CAGGGCGGTG ATCTCCTCTT	4380
CGCCGATCGT CTGGCTCAGG CGGGTCAGGC CCCGCGCATC GCTGGCCAGT TCAGCCCCCA	4440
GCACGAACAG CGTCTGCTGA ATATGGTGCA GGCTTTCCCG CAGCCCGCGC TCGCGGGTCG	4500
TGGCGTAGCA GACGCCCAGC TGGGATATCA GTTCATCGAC GGTGCCGTAG GCCTCGACGC	4560
GAATATGGTC TTTCTCGATG CGGCTGCCGC CGTACAGGGC GGTGGTGCCT TTATCCCCGG	4620
TGCGGGTATA GATACGATAC ATTCACTTTC TCTCACTTAA CGGCAGGACT TTAACCAGCT	4680
GCCCCGCGTT GCGCGCCGAG GTACGCAGTT GATCGTCGCT ATCGGTGACG TGTCCGGTAG	4740
CCAGCGGCGC GTCCGCCGGC AGCTGGGCAT GAGTGAGGGC TATCTCGCCG GACGCGCTGA	4800
GCCCCGATACC CACCCGAGG GCGGAGCTTC TGGCCGCCAG GCGGCCAGC GCAGCGGCGT	4860
CACCGCCTCC GTCATAGGTT ATGGTCTGGC AGGGGACCCC CTGCTCCTCC AGCCCCCAGC	4920
ACAGCTCATT GATGGCGCCG GCATGGTGCC CGCGCGGATC GTAAAACAGG CGTACGCCTG	4980
GCGGTGAAAG CGACATGACG GTCCCTCGT TAACACTCAG AATGCCTGGC GGAAAATCGC	5040
GGCAATCTCC TGCTCGTTGC CTTTACGCGG GTTCGAGAAC GCATTGCCGT CTTTLAGAGC	5100
CATCTCCGCC ATGTAGGGGA AGTCGGCCTC TTTTACCCC AGATCGCGCA GATGCTGCGG	5160
AATACCGATA TCCATCGACA GACGCGTGAT AGCGGCGATG GCTTTTCCG CCGCGTCGAG	5220
AGTGACAGT CCGGTGATAT TTTGCCCCAT CAGTTCAGCG ATATCGGCCA ATTTCTCCGG	5280
GTTGGCGATC AGGTTGTAGC GCGCCACATG CGGCAGCAGG ACAGCGTTGG CCACGCCGTG	5340
CGGCATGTCG TACAGGCCGC CCAGCTGGTG CGCCATGGCG TGCACGTAGC CGAGGTTGGC	5400
GTTATTGAAA GCCATCCCGG CCAGCAGAGA AGCATAGGCC ATGTTTTCCC GCGCCTGCAG	5460
ATTGCTGCCG AGGGCCACGG CCTGGCGCAG GTTGCGGGCG ATGAGGCGGA TCGCCTGCAT	5520
GGCGGCGGCG TCCGTCACCG GGTAGCGTC TTTGGAGATA TAGGCCTCTA CGGCGTGGGT	5580
CAGGGCATCC ATCCCGGTCTG CCGCGGTCAG GCGGCGCGGT TTACCGATCA TCAGCAGTGG	5640

ATCGTTGATA	GAGACCGACG	GCAGTTTGCG	CCAGCTGACG	ATCACAAACT	TCACTTTGGT	5700
TTCGGTGTG	GTCAGGACGC	AGTGGCGGGT	GACCTCGCTG	GCGGTGCCGG	CGGTGGTATT	5760
GACCGCGACG	ATAGGCGGCA	GCGGGTTGGT	CAGGGTCTCG	ATTCCGGCAT	ACTGGTACAG	5820
ATCGCCCTCA	TGGGTGGCGG	CGATGCCGAT	GCCTTTGCCG	CAATCGTGCG	GGCTGCCGCC	5880
GCCCACGGTG	ACGATGATGT	CGCACTGTTC	GCGGCGAAAC	ACGGCGAGGC	CGTCGCGCAC	5940
GTTGGTGTCT	TTCGGGTTCG	GCTCGACGCC	GTCAAAGATC	GCCACCTCGA	TCCCGGCCTC	6000
CCGCAGATAA	TGCAGGGTTT	TGTCCACCGC	GCCATCTTTA	ATTGCCCCGA	GGCCTTTGTC	6060
GGTGACCAGC	AGGGCTTTTT	TCCCCCCCAG	CAGCTGGCAG	CGTTCGCCGA	CTACGAAAT	6120
GGCGTTGGGG	CCAAAAAAGT	TAACGTTTGG	CACCAGATAA	TCAAACATAC	GATAGCTCAT	6180
AATATACCTT	CTCGCTTCAG	GTTATAATGC	GGAAAAACAA	TCCAGGGCGC	ACTGGGCTAA	6240
TAATTGATCC	TGCTCGACCG	TACCGCCGCT	AACGCCGACG	GCGCCAATTA	CCTGCTCATT	6300
AAAAATAACT	GGCAGGCCGC	CGCCAAAAAT	AATAATTCGC	TGTTGGTTGG	TTAGCTGCAG	6360
ACCGTACAGA	GATTGTCTTG	GCTGGACCGC	TGACGTAATT	TCATGGGTAC	CTTGCTTCAG	6420
GCTGCAGGCG	CTCCAGGCTT	TATTCAGGGA	AATATCGCAG	CTGGAGACGA	AGGCCTCGTC	6480
CATCCGCTGG	ATAAGCAGCG	TGTTGCCTCC	GCGGTCAACT	ACGGAACA	CCACCGCCAC	6540
GTTGATCTCA	GTGGCTTTTT	TTTCCACCGC	CGCCGCCATT	TGCTGGGCGG	CGGCCAGGGT	6600
GATTGTCTGA	ACTTGTTGGC	TCTTGTTTCT	CATTCTCTCC	CGCACCAGGA	TAACGCTGGC	6660
GCGAATAGTC	AGTAGGGGGC	GATAGTAAAA	AACTATTACC	ATTGCGTTGG	CTTGCTTTAT	6720
TTTTGTGAGC	GTTATTTTGT	CGCCGCCCAT	GATTTAGTCA	ATAGGGTTAA	AATAGCGTCG	6780
GAAAAACGTA	ATTAAGGGCG	TTTTTTATTA	ATTGATTTAT	ATCATTGCGG	GCGATCACAT	6840
TTTTTATTTT	TGCCGCCGGA	GTAAAGTTTC	ATAGTGAAAC	TGTCGGTAGA	TTTCGTGTGC	6900
CAAATTGAAA	CGAAATTAAA	TTTATTTTTT	TCACCACTGG	CTCATTTAAA	GTTCCGCTAT	6960
TGCCGGTAAT	GGCCGGGCGG	CAACGACGCT	GGCCGGCGT	ATTGCTACC	GTCTGCGGAT	7020
TTCACCTTTT	GAGCCGATGA	ACAATGAAAA	GATCAAAACG	ATTTGCAGTA	CTGGCCGAGC	7080
GCCCCGTCAA	TCAGGACGGG	CTGATTGGCG	AGTGGCCTGA	AGAGGGGCTG	ATCGCCATGG	7140
ACAGCCCCTT	TGACCCGGTC	TCTTCAGTAA	AAGTGGACAA	CGGTCTGATC	GTCGAACTGG	7200
ACGGCAAACG	CCGGGACCAG	TTTGACATGA	TCGACCGATT	TATCGCCGAT	TACGCGATCA	7260
ACGTTGAGCG	CACAGAGCAG	GCAATGCGCC	TGGAGGCGGT	GGAAATAGCC	CGTATGCTGG	7320
TGGATATTCA	CGTCAGCCGG	GAGGAGATCA	TTGCCATCAC	TACCGCCATC	ACGCCGGCCA	7380
AAGCGGTCGA	GGTGATGGCG	CAGATGAACG	TGGTGGAGAT	GATGATGGCG	CTGCAGAAGA	7440
TGCGTGCCCC	CCGGACCCCC	TCCAACCACT	GCCACGTCAC	CAATCTCAAA	GATAATCCGG	7500
TGCAGATTGC	CGCTGACGCC	GCCGAGGCCG	GGATCCGCGG	CTTCTCAGAA	CAGGAGACCA	7560
CGGTGCGTAT	CGCGCGCTAC	GCGCCGTTTA	ACGCCCTGGC	GCTGTTGGTC	GGTTCGCAGT	7620

GCGGCCGCCC	CGGCGTGTTG	ACGCAGTGCT	CGGTGGAAGA	GGCCACCGAG	CTGGAGCTGG	7680
GCGTGCCTGG	CTTAACCAGC	TACGCCGAGA	CGGTGTCGGT	CTACGGCACC	GAAGCGGTAT	7740
TTACCGACGG	CGATGATACG	CCGTGGTCAA	AGGCGTTCCT	CGCCTCGGCC	TACGCCTCCC	7800
GCGGGTTGAA	AATGCGCTAC	ACCTCCGGCA	CCGGATCCGA	AGCGCTGATG	GGCTATTCCG	7860
AGAGCAAGTC	GATGCTCTAC	CTCGAATCGC	GCTGCATCTT	CATTACTAAA	GGCGCCGGGG	7920
TTCAGGGACT	GCAAAACGGC	GCGGTGAGCT	GTATCGGCAT	GACCGGCGCT	GTGCCGTCGG	7980
GCAATCGGGC	GGTGCTGGCG	GAAAACCTGA	TCGCCTCTAT	GCTCGACCTC	GAAGTGGCGT	8040
CCGCCAACGA	CCAGACTTTC	TCCCACTCGG	ATATTGCCCG	CACCGCGCGC	ACCCTGATGC	8100
AGATGCTGCC	GGGCACCGAC	TTTATTTTCT	CCGGCTACAG	CGCGGTGCCG	AACTACGACA	8160
ACATGTTCGC	CGGCTCGAAC	TCGATGCGG	AAGATTTTGA	TGATTACAAC	ATCCTGCAGC	8220
GTGACCTGAT	GGTTGACGGC	GGCCTGCGTC	CGGTGACCGA	GGCGGAAACC	ATTGCCATTC	8280
GCCAGAAAGC	GGCGCGGGCG	ATCCAGGCGG	TTTTCCGCGA	GCTGGGGCTG	CCGCCAATCG	8340
CCGACGAGGA	GGTGGAGGCC	GCCACCTACG	CGCACGGCAG	CAACGAGATG	CCGCCGCGTA	8400
ACGTGGTGGA	GGATCTGAGT	GCGGTGGAAG	AGATGATGAA	GCGCAACATC	ACCGGCCTCG	8460
ATATTGTCGG	CGCGCTGAGC	CGCAGCGGCT	TTGAGGATAT	CGCCAGCAAT	ATTCTCAATA	8520
TGCTGCGCCA	GCGGGTCACC	GGCGATTACC	TGCAGACCTC	GGCCATTCTC	GATCGGCAGT	8580
TCGAGGTGGT	GAGTGCGGTC	AACGACATCA	ATGACTATCA	GGGGCCGGGC	ACCGGCTATC	8640
GCACTCTGCG	CGAACGCTGG	GCGGAGATCA	AAAATATTCC	GGGCGTGGTT	CAGCCCAGACA	8700
CCATTGAATA	AGGCGGTATT	CCTGTGCAAC	AGACAACCCA	AATTCAGCCC	TCTTTTACCC	8760
TGAAAACCCG	CGAGGGCGGG	GTAGCTTCTG	CCGATGAACG	CGCCGATGAA	GTGGTGATCG	8820
GCGTCGGCCC	TGCCTTCGAT	AAACACCAGC	ATCACACTCT	GATCGATATG	CCCCATGGCG	8880
CGATCCTCAA	AGAGCTGATT	GCCGGGGTGG	AAGAAGAGGG	GCTTCACGCC	CGGGTGGTGC	8940
GCAATCTGCG	CACGTCCGAC	GTCTCCTTTA	TGGCCTGGGA	TGCGGCCAAC	CTGAGCGGCT	9000
CGGGGATCGG	CATCGGTATC	CAGTCGAAGG	GGACCACGGT	CATCCATCAG	CGCGATCTGC	9060
TGCCGCTCAG	CAACCTGGAG	CTGTTCTCCC	AGGCGCCGCT	GCTGACGCTG	GAGACCTACC	9120
GGCAGATTGG	CAAAAACGCT	GCGCGCTATG	CGCGCAAAGA	GTCACCTTCG	CCGGTGCCGG	9180
TGGTGAACGA	TCAGATGGTG	CGGCCGAAAT	TTATGGCCAA	AGCCGCGCTA	TTTCATATCA	9240
AAGAGACCAA	ACATGTGGTG	CAGGACGCCG	AGCCCGTCAC	CCTGCACATC	GACTTAGTAA	9300
GGGAGTGACC	ATGAGCGAGA	AAACCATGCG	CGTGCAGGAT	TATCCGTTAG	CCACCCGCTG	9360
CCCGGAGCAT	ATCCTGACGC	CTACCGGCAA	ACCATTGACC	GATATTACCC	TCGAGAAGGT	9420
GCTCTCTGGC	GAGGTGGGCC	CGCAGGATGT	GCGGATCTCC	CGCCAGACCC	TTGAGTACCA	9480
GGCGCAGATT	GCCGAGCAGA	TGCAGCGCCA	TGCGGTGGCG	CGCAATTTCC	GCCGCGCGGC	9540
GGAGCTTATC	GCCATTCTCTG	ACGAGCGCAT	TCTGGCTATC	TATAACGCGC	TGCGCCCGTT	9600

CCGCTCCTCG CAGGCGGAGC TGCTGGCGAT CGCCGACGAG CTGGAGCACA CCTGGCATGC 9660
 GACAGTGAAT GCCGCCTTTG TCCGGGAGTC GGCGBAAGTG TATCAGCAGC GGCATAAGCT 9720
 GCGTAAAGGA AGCTAAGCGG AGGTCAGCAT GCCGTTAATA GCCGGGATTG ATATCGGCAA 9780
 CGCCACCACC GAGGTGGCGC TGGCGTCCGA CTACCCGCAG GCGAGGGCGT TTGTGTCCAG 9840
 CGGGATCGTC GCGACGACGG GCATGAAAGG GACGCGGGAC AATATCGCCG GGACCCTCGC 9900
 CGCGCTGGAG CAGGCCCTGG CGAAAACACC GTGGTCGATG AGCGATGTCT CTCGCATCTA 9960
 TCTTAACGAA GCCGCGCCGG TGATTGGCGA TGTGGCGATG GAGACCATCA CCGAGACCAT 10020
 TATCACCGAA TCGACCATGA TCGGTCATAA CCCGCAGACG CCGGGCGGGG TGGGCGTTGG 10080
 CGTGGGGACG ACTATCGCCC TCGGGCGGCT GCGGACGCTG CCGGGCGGCG AGTATGCCGA 10140
 GGGGTGGATC GTACTGATTG ACGACGCCGT CGATTTCTT GACGCCGTGT GGTGGCTCAA 10200
 TGAGGCGCTC GACCGGGGGA TCAACGTGGT GCGGCGATC CTCAAAAAGG ACGACGGCGT 10260
 GCTGGTGAAC AACC GCCTGC GTAAAACCCT GCCGGTGGTG GATGAAGTGA CGCTGCTGGA 10320
 GCAGGTCCCC GAGGGGGTAA TGGCGGCGGT GGAAGTGGCC GCGCCGGGCC AGGTGGTGCG 10380
 GATCCTGTCT AATCCCTACG GGATCGCCAC CTTCTTCGGG CTAAGCCCGG AAGAGACCCA 10440
 GGCCATCGTC CCCATCGCCC GCGCCCTGAT TGGAACCGT TCCGCGGTGG TGCTCAAGAC 10500
 CCCGCAGGGG GATGTGCACT GCGGGGTGAT CCCGGCGGGC AACCTCTACA TTAGCGGCGA 10560
 AAAGCGCCGC GGAGAGGCCG ATGTCGCCGA GGGCGCGGAA GCCATCATGC AGGCGATGAG 10620
 CGCCTGCGCT CCGGTACGCG ACATCCGCGG CGAACCGGGC ACCCACGCCG GCGGCATGCT 10680
 TGAGCGGGTG CGCAAGGTAA TGGCGTCCCT GACCGGCCAT GAGATGAGCG CGATATACAT 10740
 CCAGGATCTG CTGGCGGTGG ATACGTTTAT TCCGCGCAAG GTGCAGGGCG GGATGGCCGG 10800
 CGAGTGCGCC ATGGAGAATG CCGTCGGGAT GCGGCGGATG GTGAAAGCGG ATCGTCTGCA 10860
 AATGCAGGTT ATCGCCCGCG AACTGAGCGC CCGACTGCAG ACCGAGGTGG TGGTGGGCGG 10920
 CGTGGAGGCC AACATGGCCA TCGCCGGGGC GTTAACCACT CCCGGCTGTG CGGCGCCGCT 10980
 GCGGATCCTC GACCTCGGCG CCGGCTCGAC GGATGCGGCG ATCGTCAACG CGGAGGGGCA 11040
 GATAACGGCG GTCCATCTCG CCGGGGCGGG GAATATGGTC AGCCTGTTGA TTAAACCGA 11100
 GCTGGGCCTC GAGGATCTTT CGCTGGCGGA AGCGATAAAA AAATACCCGC TGGCCAAAGT 11160
 GGAAAGCCTG TTCAGTATTC GTCACGAGAA TGGCGCGGTG GAGTTCTTTC GGAAGCCCT 11220
 CAGCCCGGCG GTGTTGCCA AAGTGGTGTA CATCAAGGAG GGCGAACTGG TGCCGATCGA 11280
 TAACGCCAGC CCGCTGGAAA AAATTCGTCT CGTGCGCCG CAGGCGAAAG AGAAAGTGTT 11340
 TGTCACCAAC TGCTGCGCG CGCTGCGCCA GGTCTCAGCC GCGGTTCCA TTCGCGATAT 11400
 CGCCTTTGTG GTGCTGGTGG GCGGCTCATC GCTGGACTTT GAGATCCCGC AGCTTATCAC 11460
 GGAAGCCTTG TCGCACTATG GCGTGGTCGC CGGGCAGGGC AATATTCGGG GAACAGAAGG 11520
 GCCGCGCAAT GCGGTGCGCA CCGGGCTGCT ACTGGCCGGT CAGGCGAATT AAACGGGCGC 11580

TCGCGCCAGC CTCTCTCTTT AACGTGCTAT TTCAGGATGC CGATAATGAA CCAGACTTCT 11640
 ACCTTAACCG GGCAGTGCCT GGCCGAGTTT CTTGGCACCG GATTGCTCAT TTTCTTCGGC 11700
 GCGGGCTGCG TCGCTGCGCT GCGGGTCGCC GGGGCCAGCT TTGGTCAGTG GGAGATCAGT 11760
 ATTATCTGGG GCCTTGGCGT CGCCATGGCC ATCTACCTGA CGGCCGGTGT CTCCGGCGCG 11820
 CACCTAAATC CGGCGGTGAC CATTGCCCTG TGGCTGTTCG CCTGTTTTGA ACGCCGCAAG 11880
 GTGCTGCCGT TTATGTGTC CCAGACGGCC GGGGCCTTCT GCGCCGCCGC GCTGGTGTAT 11940
 GGGCTCTATC GCCAGCTGTT TCTCGATCTT GAACAGAGTC AGCATATCGT GCGCGGCACT 12000
 GCCGCCAGTC TTAACCTGGC CGGGGTCTTT TCCACGTACC CGCATCCACA TATCACTTTT 12060
 ATACAAGCGT TTGCCGTGGA GACCACCATC ACGGCAATCC TGATGGCGAT GATCATGGCC 12120
 CTGACCGACG ACGGCAACGG AATTC 12145

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 94 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

AGCTTAGGAG TCTAGAATAT TGAGCTCGAA TTCCCGGGCA TCGGTACCG GATCCAGAAA 60
 AAAGCCCGCA CCTGACAGTG CGGGCTTTTT TTTT 94

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 37 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GGAATTCAGA TCTCAGCAAT GAGCGAGAAA ACCATGC 37

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 27 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GCTCTAGATT AGCTTCCTTT ACGCAGC 27

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:
GGCCAAGCTT AAGGAGGTTA ATTAAATGAA AAG 33
- (2) INFORMATION FOR SEQ ID NO:24:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:
GCTCTAGATT ATTCAATGGT GTCGGG 26
- (2) INFORMATION FOR SEQ ID NO:25:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:
GCGCCGTCTA GAATTATGAG CTATCGTATG TTTGATTATC TG 42
- (2) INFORMATION FOR SEQ ID NO:26:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:
TCTGATACGG GATCCTCAGA ATGCCTGGCG GAAAAT 36
- (2) INFORMATION FOR SEQ ID NO:27:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 51 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:
GCGCGGATCC AGGAGTCTAG AATTATGGGA TTGACTACTA AACCTCTATC T 51
- (2) INFORMATION FOR SEQ ID NO:28:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:
GATACGCCCG GGTTACCATT TCAACAGATC GTCCTT 36
- (2) INFORMATION FOR SEQ ID NO:29:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:
TCGACGAATT CAGGAGGA 18
- (2) INFORMATION FOR SEQ ID NO:30:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:
CTAGTCCTCC TGAATTCC 18
- (2) INFORMATION FOR SEQ ID NO:31:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:
CTAGTAAGGA GGACAATTC 19
- (2) INFORMATION FOR SEQ ID NO:32:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

CATGGAATTG TCCTCCTTA

19

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 271 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: GPPI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Met	Lys	Arg	Phe	Asn	Val	Leu	Lys	Tyr	Ile	Arg	Thr	Thr	Lys	Ala	Asn	1	5	10	15
Ile	Gln	Thr	Ile	Ala	Met	Pro	Leu	Thr	Thr	Lys	Pro	Leu	Ser	Leu	Lys	20	25	30	
Ile	Asn	Ala	Ala	Leu	Phe	Asp	Val	Asp	Gly	Thr	Ile	Ile	Ile	Ser	Gln	35	40	45	
Pro	Ala	Ile	Ala	Ala	Phe	Trp	Arg	Asp	Phe	Gly	Lys	Asp	Lys	Pro	Tyr	50	55	60	
Phe	Asp	Ala	Glu	His	Val	Ile	His	Ile	Ser	His	Gly	Trp	Arg	Thr	Tyr	65	70	75	80
Asp	Ala	Ile	Ala	Lys	Phe	Ala	Pro	Asp	Phe	Ala	Asp	Glu	Glu	Tyr	Val	85	90	95	
Asn	Lys	Leu	Glu	Gly	Glu	Ile	Pro	Glu	Lys	Tyr	Gly	Glu	His	Ser	Ile	100	105	110	
Glu	Val	Pro	Gly	Ala	Val	Lys	Leu	Cys	Asn	Ala	Leu	Asn	Ala	Leu	Pro	115	120	125	
Lys	Glu	Lys	Trp	Ala	Val	Ala	Thr	Ser	Gly	Thr	Arg	Asp	Met	Ala	Lys	130	135	140	
Lys	Trp	Phe	Asp	Ile	Leu	Lys	Ile	Lys	Arg	Pro	Glu	Tyr	Phe	Ile	Thr	145	150	155	160
Ala	Asn	Asp	Val	Lys	Gln	Gly	Lys	Pro	His	Pro	Glu	Pro	Tyr	Leu	Lys	165	170	175	
Gly	Arg	Asn	Gly	Leu	Gly	Phe	Pro	Ile	Asn	Glu	Gln	Asp	Pro	Ser	Lys	180	185	190	
Ser	Lys	Val	Val	Val	Phe	Glu	Asp	Ala	Pro	Ala	Gly	Ile	Ala	Ala	Gly	195	200	205	
Lys	Ala	Ala	Gly	Cys	Lys	Ile	Val	Gly	Ile	Ala	Thr	Thr	Phe	Asp	Leu	210	215	220	

Asp Phe Leu Lys Glu Lys Gly Cys Asp Ile Ile Val Lys Asn His Glu
 225 230 235 240
 Ser Ile Arg Val Gly Glu Tyr Asn Ala Glu Thr Asp Glu Val Glu Leu
 245 250 255
 Ile Phe Asp Asp Tyr Leu Tyr Ala Lys Asp Asp Leu Leu Lys Trp
 260 265 270

(2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 555 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: unknown
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: DHAB1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Met Lys Arg Ser Lys Arg Phe Ala Val Leu Ala Gln Arg Pro Val Asn
 1 5 10 15
 Gln Asp Gly Leu Ile Gly Glu Trp Pro Glu Glu Gly Leu Ile Ala Met
 20 25 30
 Asp Ser Pro Phe Asp Pro Val Ser Ser Val Lys Val Asp Asn Gly Leu
 35 40 45
 Ile Val Glu Leu Asp Gly Lys Arg Arg Asp Gln Phe Asp Met Ile Asp
 50 55 60
 Arg Phe Ile Ala Asp Tyr Ala Ile Asn Val Glu Arg Thr Glu Gln Ala
 65 70 75 80
 Met Arg Leu Glu Ala Val Glu Ile Ala Arg Met Leu Val Asp Ile His
 85 90 95
 Val Ser Arg Glu Glu Ile Ile Ala Ile Thr Thr Ala Ile Thr Pro Ala
 100 105 110
 Lys Ala Val Glu Val Met Ala Gln Met Asn Val Val Glu Met Met Met
 115 120 125
 Ala Leu Gln Lys Met Arg Ala Arg Arg Thr Pro Ser Asn Gln Cys His
 130 135 140
 Val Thr Asn Leu Lys Asp Asn Pro Val Gln Ile Ala Ala Asp Ala Ala
 145 150 155 160
 Glu Ala Gly Ile Arg Gly Phe Ser Glu Gln Glu Thr Thr Val Gly Ile
 165 170 175
 Ala Arg Tyr Ala Pro Phe Asn Ala Leu Ala Leu Leu Val Gly Ser Gln
 180 185 190
 Cys Gly Arg Pro Gly Val Leu Thr Gln Cys Ser Val Glu Glu Ala Thr
 195 200 205
 Glu Leu Glu Leu Gly Met Arg Gly Leu Thr Ser Tyr Ala Glu Thr Val
 210 215 220

Ser Val Tyr Gly Thr Glu Ala Val Phe Thr Asp Gly Asp Asp Thr Pro
 225 230 235 240
 Trp Ser Lys Ala Phe Leu Ala Ser Ala Tyr Ala Ser Arg Gly Leu Lys
 245 250 255
 Met Arg Tyr Thr Ser Gly Thr Gly Ser Glu Ala Leu Met Gly Tyr Ser
 260 265 270
 Glu Ser Lys Ser Met Leu Tyr Leu Glu Ser Arg Cys Ile Phe Ile Thr
 275 280 285
 Lys Gly Ala Gly Val Gln Gly Leu Gln Asn Gly Ala Val Ser Cys Ile
 290 295 300
 Gly Met Thr Gly Ala Val Pro Ser Gly Ile Arg Ala Val Leu Ala Glu
 305 310 315 320
 Asn Leu Ile Ala Ser Met Leu Asp Leu Glu Val Ala Ser Ala Asn Asp
 325 330 335
 Gln Thr Phe Ser His Ser Asp Ile Arg Arg Thr Ala Arg Thr Leu Met
 340 345 350
 Gln Met Leu Pro Gly Thr Asp Phe Ile Phe Ser Gly Tyr Ser Ala Val
 355 360 365
 Pro Asn Tyr Asp Asn Met Phe Ala Gly Ser Asn Phe Asp Ala Glu Asp
 370 375 380
 Phe Asp Asp Tyr Asn Ile Leu Gln Arg Asp Leu Met Val Asp Gly Gly
 385 390 395 400
 Leu Arg Pro Val Thr Glu Ala Glu Thr Ile Ala Ile Arg Gln Lys Ala
 405 410 415
 Ala Arg Ala Ile Gln Ala Val Phe Arg Glu Leu Gly Leu Pro Pro Ile
 420 425 430
 Ala Asp Glu Glu Val Glu Ala Ala Thr Tyr Ala His Gly Ser Asn Glu
 435 440 445
 Met Pro Pro Arg Asn Val Val Glu Asp Leu Ser Ala Val Glu Glu Met
 450 455 460
 Met Lys Arg Asn Ile Thr Gly Leu Asp Ile Val Gly Ala Leu Ser Arg
 465 470 475 480
 Ser Gly Phe Glu Asp Ile Ala Ser Asn Ile Leu Asn Met Leu Arg Gln
 485 490 495
 Arg Val Thr Gly Asp Tyr Leu Gln Thr Ser Ala Ile Leu Asp Arg Gln
 500 505 510
 Phe Glu Val Val Ser Ala Val Asn Asp Ile Asn Asp Tyr Gln Gly Pro
 515 520 525
 Gly Thr Gly Tyr Arg Ile Ser Ala Glu Arg Trp Ala Glu Ile Lys Asn
 530 535 540
 Ile Pro Gly Val Val Gln Pro Asp Thr Ile Glu
 545 550 555

(2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 194 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: unknown
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: DHAB2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

```

Met Gln Gln Thr Thr Gln Ile Gln Pro Ser Phe Thr Leu Lys Thr Arg
 1             5             10             15

Glu Gly Gly Val Ala Ser Ala Asp Glu Arg Ala Asp Glu Val Val Ile
      20             25             30

Gly Val Gly Pro Ala Phe Asp Lys His Gln His His Thr Leu Ile Asp
      35             40             45

Met Pro His Gly Ala Ile Leu Lys Glu Leu Ile Ala Gly Val Glu Glu
      50             55             60

Glu Gly Leu His Ala Arg Val Val Arg Ile Leu Arg Thr Ser Asp Val
      65             70             75             80

Ser Phe Met Ala Trp Asp Ala Ala Asn Leu Ser Gly Ser Gly Ile Gly
      85             90             95

Ile Gly Ile Gln Ser Lys Gly Thr Thr Val Ile His Gln Arg Asp Leu
      100            105            110

Leu Pro Leu Ser Asn Leu Glu Leu Phe Ser Gln Ala Pro Leu Leu Thr
      115            120            125

Leu Glu Thr Tyr Arg Gln Ile Gly Lys Asn Ala Ala Arg Tyr Ala Arg
      130            135            140

Lys Glu Ser Pro Ser Pro Val Pro Val Val Asn Asp Gln Met Val Arg
      145            150            155            160

Pro Lys Phe Met Ala Lys Ala Ala Leu Phe His Ile Lys Glu Thr Lys
      165            170            175

His Val Val Gln Asp Ala Glu Pro Val Thr Leu His Ile Asp Leu Val
      180            185            190

Arg Glu

```

(2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 140 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: unknown
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: DHAB3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Met Ser Glu Lys Thr Met Arg Val Gln Asp Tyr Pro Leu Ala Thr Arg
 1 5 10 15
 Cys Pro Glu His Ile Leu Thr Pro Thr Gly Lys Pro Leu Thr Asp Ile
 20 25 30
 Thr Leu Glu Lys Val Leu Ser Gly Glu Val Gly Pro Gln Asp Val Arg
 35 40 45
 Ile Ser Arg Gln Thr Leu Glu Tyr Gln Ala Gln Ile Ala Glu Gln Met
 50 55 60
 Gln His Ala Val Ala Arg Asn Phe Arg Arg Ala Ala Glu Leu Ile Ala
 65 70 75 80
 Ile Pro Asp Glu Arg Ile Leu Ala Ile Tyr Asn Ala Leu Arg Pro Phe
 85 90 95
 Arg Ser Ser Gln Ala Glu Leu Leu Ala Ile Ala Asp Glu Leu Glu His
 100 105 110
 Thr Trp His Ala Thr Val Asn Ala Ala Phe Val Arg Glu Ser Ala Glu
 115 120 125
 Val Tyr Gln Gln Arg His Lys Leu Arg Lys Gly Ser
 130 135 140

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 387 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: unknown
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: DHAT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Met Ser Tyr Arg Met Phe Asp Tyr Leu Val Pro Asn Val Asn Phe Phe
 1 5 10 15
 Gly Pro Asn Ala Ile Ser Val Val Gly Glu Arg Cys Gln Leu Leu Gly
 20 25 30
 Gly Lys Lys Ala Leu Leu Val Thr Asp Lys Gly Leu Arg Ala Ile Lys
 35 40 45
 Asp Gly Ala Val Asp Lys Thr Leu His Tyr Leu Arg Glu Ala Gly Ile
 50 55 60
 Glu Val Ala Ile Phe Asp Gly Val Glu Pro Asn Pro Lys Asp Thr Asn
 65 70 75 80
 Val Arg Asp Gly Leu Ala Val Phe Arg Arg Glu Gln Cys Asp Ile Ile
 85 90 95
 Val Thr Val Gly Gly Gly Ser Pro His Asp Cys Gly Lys Gly Ile Gly
 100 105 110

```

Ile Ala Ala Thr His Glu Gly Asp Leu Tyr Gln Tyr Ala Gly Ile Glu
    115                      120                      125

Thr Leu Thr Asn Pro Leu Pro Pro Ile Val Ala Val Asn Thr Thr Ala
    130                      135                      140

Gly Thr Ala Ser Glu Val Thr Arg His Cys Val Leu Thr Asn Thr Glu
    145                      150                      155                      160

Thr Lys Val Lys Phe Val Ile Val Ser Trp Arg Lys Leu Pro Ser Val
    165                      170                      175

Ser Ile Asn Asp Pro Leu Leu Met Ile Gly Lys Pro Ala Ala Leu Thr
    180                      185                      190

Ala Ala Thr Gly Met Asp Ala Leu Thr His Ala Val Glu Ala Tyr Ile
    195                      200                      205

Ser Lys Asp Ala Asn Pro Val Thr Asp Ala Ala Ala Met Gln Ala Ile
    210                      215                      220

Arg Leu Ile Ala Arg Asn Leu Arg Gln Ala Val Ala Leu Gly Ser Asn
    225                      230                      235                      240

Leu Gln Ala Arg Glu Asn Met Ala Tyr Ala Ser Leu Leu Ala Gly Met
    245                      250                      255

Ala Phe Asn Asn Ala Asn Leu Gly Tyr Val His Ala Met Ala His Gln
    260                      265                      270

Leu Gly Gly Leu Tyr Asp Met Pro His Gly Val Ala Asn Ala Val Leu
    275                      280                      285

Leu Pro His Val Ala Arg Tyr Asn Leu Ile Ala Asn Pro Glu Lys Phe
    290                      295                      300

Ala Asp Ile Ala Glu Leu Met Gly Glu Asn Ile Thr Gly Leu Ser Thr
    305                      310                      315                      320

Leu Asp Ala Ala Glu Lys Ala Ile Ala Ala Ile Thr Arg Leu Ser Met
    325                      330                      335

Asp Ile Gly Ile Pro Gln His Leu Arg Asp Leu Gly Val Lys Glu Ala
    340                      345                      350

Asp Phe Pro Tyr Met Ala Glu Met Ala Leu Lys Asp Gly Asn Ala Phe
    355                      360                      365

Ser Asn Pro Arg Lys Gly Asn Glu Gln Glu Ile Ala Ala Ile Phe Arg
    370                      375                      380

Gln Ala Phe
    385

```

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:
GCGAATTCAT GAGCTATCGT ATGTTTG 27
- (2) INFORMATION FOR SEQ ID NO:39:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:
GCGAATTCAG AATGCCTGGC GGAAAATC 28
- (2) INFORMATION FOR SEQ ID NO:40:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:
GGGAATTCAT GAGCGAGAAA ACCATGCG 28
- (2) INFORMATION FOR SEQ ID NO:41:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:
GCGAATTCTT AGCTTCCTTT ACGCAGC 27
- (2) INFORMATION FOR SEQ ID NO:42:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:
GCGAATTCAT GCAACAGACA ACCCAAATC 30
- (2) INFORMATION FOR SEQ ID NO:43:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:
- GCGAATTCAC TCCCTTACTA AGTCG 25
- (2) INFORMATION FOR SEQ ID NO:44:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:
- GGGAATTCAT GAAAAGATCA AAACGATTTG 30
- (2) INFORMATION FOR SEQ ID NO:45:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:
- GCGAATTCTT ATTCAATGGT GTCGGGCTG 29
- (2) INFORMATION FOR SEQ ID NO:46
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:
- TTGATAATAT AACCATGGCT GCTGCTGCTG ATAG 34
- (2) INFORMATION FOR SEQ ID NO:47
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:
- GTATGATATG TTATCTTGGA TCCAATAAAT CTAATCTTC 39

(2) INFORMATION FOR SEQ ID NO:48:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

CATGACTAGT AAGGAGGACA ATTC

24

(2) INFORMATION FOR SEQ ID NO:49:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

CATGGAATTG TCCTCCTTAC TAGT

24

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>7 and 8</u> , lines <u>37 & 38</u> on pg. 7 & Lines 1-5 on pg. 8	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution AMERICAN TYPE CULTURE COLLECTION	
Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 US	
Date of deposit 26 September 1996	Accession Number 98188
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
In respect of those designations in which a European patent is sought, a sample of the deposited microorganism will be made available until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample. (Rule 28(4) EPC)	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")	

<div style="border-bottom: 1px solid black; padding-bottom: 5px;"> For receiving Office use only </div> <div style="padding: 5px;"> <input type="checkbox"/> This sheet was received with the international application </div> <div style="border-top: 1px solid black; padding-top: 5px;"> Authorized officer </div>	<div style="border-bottom: 1px solid black; padding-bottom: 5px;"> For International Bureau use only </div> <div style="padding: 5px;"> <input type="checkbox"/> This sheet was received by the International Bureau on: </div> <div style="border-top: 1px solid black; padding-top: 5px;"> Authorized officer </div>
--	---

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>8</u> , line s <u>6 - 12</u>	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution AMERICAN TYPE CULTURE COLLECTION	
Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 US	
Date of deposit 26 September 1996	Accession Number 74392
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
In respect of those designations in which a European patent is sought, a sample of the deposited microorganism will be made available until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample. (Rule 28(4) EPC)	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")	

<div style="border-bottom: 1px solid black; padding-bottom: 5px;"> For receiving Office use only </div> <div style="padding: 5px;"> <input type="checkbox"/> This sheet was received with the international application </div> <div style="border-top: 1px solid black; padding-top: 5px;"> Authorized officer </div>	<div style="border-bottom: 1px solid black; padding-bottom: 5px;"> For International Bureau use only </div> <div style="padding: 5px;"> <input type="checkbox"/> This sheet was received by the International Bureau on: </div> <div style="border-top: 1px solid black; padding-top: 5px;"> Authorized officer </div>
--	---

WHAT IS CLAIMED IS:

1. A method for the production of 1,3-propanediol from a recombinant organism comprising:

- (i) transforming a suitable host organism with a transformation cassette comprising at least one of
- (a) a gene encoding a glycerol-3-phosphate dehydrogenase activity;
 - (b) a gene encoding a glycerol-3-phosphatase activity;
 - (c) genes encoding a dehydratase activity;
 - (d) a gene encoding 1,3-propanediol oxidoreductase activity,
- provided that if the transformation cassette comprises less than all the genes of (a)-(d), then the suitable host organism comprises endogenous genes whereby the resulting transformed host organism comprises at least one of each of genes (a)-(d);
- (ii) culturing the transformed host organism under suitable conditions in the presence of at least one carbon source selected from the group consisting of monosaccharides, oligosaccharides, polysaccharides, or a one-carbon substrate whereby 1,3-propanediol is produced; and
- (iii) recovering the 1,3-propanediol.

2. The method of Claim 1 wherein the transformation cassette comprises all of the genes (a)-(d).

3. The method of Claim 1 wherein the suitable host organism is selected from the group consisting of bacteria, yeast, and filamentous fungi.

4. The method of Claim 3 wherein the suitable host organism is selected from the group of genera consisting of *Citrobacter*, *Enterobacter*, *Clostridium*, *Klebsiella*, *Aerobacter*, *Lactobacillus*, *Aspergillus*, *Saccharomyces*, *Schizosaccharomyces*, *Zygosaccharomyces*, *Pichia*, *Kluyveromyces*, *Candida*, *Hansenula*, *Debaryomyces*, *Mucor*, *Torulopsis*, *Methylobacter*, *Escherichia*, *Salmonella*, *Bacillus*, *Streptomyces* and *Pseudomonas*.

5. The method of Claim 4 wherein the suitable host organism is selected from the group consisting of *E. coli*, *Klebsiella spp.*, and *Saccharomyces spp.*

6. The method of Claim 1 wherein the transformed host organism is a *Saccharomyces spp.* transformed with a transformation cassette comprising the genes *dhaB1*, *dhaB2*, *dhaB3*, and *dhaT*, wherein the genes are stably integrated into the *Saccharomyces spp.* genome.

7. The method of Claim 1 wherein the transformed host organism is a *Klebsiella spp.* transformed with a transformation cassette comprising the genes GPD1 and GPD2.

8. The method of Claim 1 wherein the carbon source is glucose.

5 9. The method of Claim 1 wherein the gene encoding a glycerol-3-phosphate dehydrogenase enzyme is selected from the group consisting of genes corresponding to amino acid sequences given in SEQ ID NO:11, in SEQ ID NO:12, and in SEQ ID NO:13, the amino acid sequences encompassing amino acid substitutions, deletions or additions that do not alter the function of the
10 glycerol-3-phosphate dehydrogenase enzyme.

10 10. The method of Claim 1 wherein the gene encoding a glycerol-3-phosphatase enzyme is selected from the group consisting of genes corresponding to amino acid sequences given in SEQ ID NO:33 and in SEQ ID NO:17, the amino acid sequences encompassing amino acid substitutions,
15 deletions or additions that do not alter the function of the glycerol-3-phosphatase enzyme.

20 11. The method of Claim 1 wherein the gene encoding a glycerol kinase enzyme corresponds to an amino acid sequence given in SEQ ID NO:18, the amino acid sequence encompassing amino acid substitutions, deletions or additions that do not alter the function of the glycerol kinase enzyme.

25 12. The method of Claim 1 wherein the genes encoding a dehydratase enzyme comprise dhaB1, dhaB2 and dhaB3, the genes corresponding respectively to amino acid sequences given in SEQ ID NO:34, SEQ ID NO:35, and SEQ ID NO:36, the amino acid sequences encompassing amino acid substitutions, deletions or additions that do not alter the function of the dehydratase enzyme.

30 13. The method of Claim 1 wherein the gene encoding a 1,3-propanediol oxidoreductase enzyme corresponds to an amino acid sequence given in SEQ ID NO:37, the amino acid sequence encompassing amino acid substitutions, deletions or additions that do not alter the function of the 1,3-propanediol oxidoreductase enzyme.

14. A transformed host cell comprising:

(a) a group of genes comprising

35 (1) a gene encoding a glycerol-3-phosphate dehydrogenase enzyme corresponding to the amino acid sequence given in SEQ ID NO:11;

(2) a gene encoding a glycerol-3-phosphatase enzyme corresponding to the amino acid sequence given in SEQ ID NO:17;

(3) a gene encoding the a subunit of the glycerol dehydratase enzyme corresponding to the amino acid sequence given in SEQ ID NO:34;

- (4) a gene encoding the β subunit of the glycerol dehydratase enzyme corresponding to the amino acid sequence given in SEQ ID NO:35;
- (5) a gene encoding the γ subunit of the glycerol dehydratase enzyme corresponding to the amino acid sequence given in SEQ ID NO:36; and
- 5 (6) a gene encoding the 1,3-propanediol oxidoreductase enzyme corresponding to the amino acid sequence given in SEQ ID NO:37, the respective amino acid sequences of (a)(1)-(6) encompassing amino acid substitutions, deletions, or additions that do not alter the function of the enzymes of genes (1)-(6), and
- 10 (b) a host cell transformed with the group of genes of (a), whereby the transformed host cell produces 1,3-propanediol on at least one substrate selected from the group consisting of monosaccharides, oligosaccharides, and polysaccharides or from a one-carbon substrate.

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 97/20292

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/53 C12N15/55 C12N15/60 C12P7/18 C12N9/04
C12N9/16 C12N9/88

According to International Patent Classification(IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X, L	WO 96 35796 A (DU PONT ; GENENCOR INT (US); LAFFEND LISA ANNE (US); NAGARAJAN VASA) 14 November 1996 see the whole document see abstract see claims 1-33 --- -/--	1, 3-6, 8, 12, 13



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

10 March 1998

Date of mailing of the international search report

24/03/1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Lejeune, R

INTERNATIONAL SEARCH REPORT

Int. l. Application No

PCT/US 97/20292

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>I-TEH TONG ET AL: "ENHANCEMENT OF 1,3-PROPANEDIOL PRODUCTION BY COFERMENTATION IN ESCHERICHIA COLI EXPRESSING KLEBSIELLA PNEUMONIAE DHA REGULON GENES" APPLIED BIOCHEMISTRY AND BIOTECHNOLOGY, vol. 34/35, 1992, pages 149-159, XP000578884 cited in the application see abstract see page 150, line 24 - line 29 see page 152; table 1</p>	1,3-6,8, 12,13
A	<p>DE 37 34 764 A (HUELS CHEMISCHE WERKE AG) 3 May 1989 cited in the application see abstract see column 3; claim 1</p>	1
A	<p>DANIEL R ET AL: "PURIFICATION OF 1,3-PROPANEDIOL DEHYDROGENASE FROM CITROBACTER FREUNDII AND CLONING, SEQUENCING, AND OVEREXPRESSION OF THE CORRESPONDING GENE IN ESCHERICHIA COLI" JOURNAL OF BACTERIOLOGY, vol. 177, no. 8, 1 April 1995, pages 2151-2156, XP000579775 cited in the application see abstract</p>	13
A	<p>TOBIMATSU T. ET AL.: "Cloning, sequencing and high level expression of the genes encoding adenosylcobalamin-dependent glycerol dehydrase of Klebsiella pneumoniae." JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 271, no. 37, 13 September 1996, pages 22352-22357, XP002057923 see abstract see page 22355; figure 5 see page 22356, column 1, line 2 - line 5</p>	12

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inte. l.ional Application No

PCT/US 97/20292

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9635796 A	14-11-96	US 5686276 A	11-11-97
		AU 5678996 A	29-11-96
		EP 0826057 A	04-03-98
<hr/>			
DE 3734764 A	03-05-89	NONE	
<hr/>			